Algorithmic Writing Assistance on Jobseekers' Resumes Increases Hires

Emma Wiles[*] MIT

Zanele Munyikwa[†] MIT

John J. Horton[‡] MIT & NBER

February 27, 2024

Most recent draft here.

Abstract

There is a strong association between writing quality in resumes for new labor market entrants and whether they are ultimately hired. We show this relationship is, at least partially, causal: in a field experiment in an online labor market with nearly half a million jobseekers, treated jobseekers received algorithmic writing assistance on their resumes. Treated jobseekers were hired 8% more often, at 10% higher wages. Contrary to concerns that the assistance takes away a valuable signal, we find no evidence that employers were less satisfied. We find that the writing on treated jobseekers resumes had fewer errors and was easier to read. Our analysis suggests that better writing is not a reliable signal of ability but helps employers *ascertain* ability through clearer writing, suggesting digital platforms could benefit from incorporating algorithmic writing assistance into text-based descriptions of labor services or products.

1 Introduction

For most employers, the first exposure to a job candidate is typically a written resume. The resume contains information about the applicant—education, skills, past employment, and so on—that the employer uses to draw inferences about the applicant's suitability for the job. A well-written resume might influence an employer's perception of a candidate. One perspective is that a better-written resume—without any change in the underlying facts—might make it easier for the employer to draw the correct inferences about a candidate's abilities, potentially improving the chance of an interview or job offer. We call this the "clarity view" of the role of resume writing quality. From another perspective, a resume might

[*]Corresponding author: emmavani@mit.edu

[†]zanele@mit.edu

[‡]john.joseph.horton@gmail.com

not merely be a conduit for match-relevant information; the resume's writing itself could signal ability. In particular, writing quality might provide signals about the jobseeker's communication skills, attention to detail, or overall quality, potentially leading to a greater chance of a positive outcome. We call this the "signaling view" of the role of resume writing quality.

In this paper, we explore the mechanics of how resume writing quality affects the hiring process. First, using observational data from a large online labor market, we document a strong positive relationship between writing quality in resumes and hiring that persists even after controlling for obvious confounders. Second, we report the results of a field experiment in which we exogenously vary writing quality in the same market. This experiment directly tests whether there is a causal effect of writing quality on job market outcomes and provides a testing ground to distinguish between the clarity and signaling views.

Our main substantive finding is evidence for the "clarity view." Evidence for this conclusion is possible because we trace the whole matching process from resume creation all the way to a measure of post-employment satisfaction with a sample of 480,948 jobseekers. This sample size is an order of magnitude larger than the next largest experiments.

Treated jobseekers were more likely to get hired (consistent with both signaling and clarity explanations), but we find no evidence that employers were later disappointed in the quality of work by the treated group, which refutes what the "signaling view" explanation would predict.

To create random variation in writing quality, we intercept new jobseekers at the resume-writing stage of registering for the online labor market. We randomly offer some of them—the treatment group—algorithmic writing assistance, while others—the control group—write their resume under the status quo experience of no assistance. We will discuss this assistance in depth, which we refer to as the Algorithmic Writing Service, but, at a high level, it improves writing by identifying and providing suggestions to resolve common errors. After resume creation, we observe both treated and control jobseekers as they engage in search and, in the case of completed jobs, receive reviews.

In the experimental data, there are quantifiable improvements to resume-writing quality among the treatment group. For example, we find fewer grammar errors, redundancies, and commonly confused words in the resumes of the treated group of jobseekers. These positive effects to writing were greatest at the low-end of the distribution in writing quality, as jobseekers with already excellent resumes benefited little from writing assistance.

One might worry that the treatment could affect behavior. However, we find that, during job search, treated workers did not send out more applications than workers in the control group, nor did they propose higher wage bids. This is a convenient result, as it allows us to focus on the decision-making of the employers. If jobseekers had altered their application behavior—perhaps sending more applications with their stronger resumes—we might wrongly attribute greater job-market success to the resume rather than this endogenous change in effort.

As for the effect of writing assistance on hiring, we find that treated jobseekers had a 8% increase in their probability of being hired within their first month on the platform relative to the control group. If hired, treated workers' hourly wages were 10% higher than the hourly wages of workers in the control group. This result is downstream of hiring and we provide evidence that it is due to changes in the composition of which workers were hired.

The data make the impact of resume quality on hiring decisions apparent. In order to differentiate between the "signaling view" and "clarity view," we look at the effect of the treatment on a few different proxy's for employers' satisfaction with the quality of work. We do not find any significant treatment effects to revealed preference measures like hours worked or whether or not workers were ever rehired.

Unique to our setting, we also have explicit measures of employer disappointment, as both sides privately rate each other at the conclusion of the contract. Employer ratings provide a direct way to analyze the informational role of the resume. Specifically, since the treatment removes or at least weakens a credible signal of jobseeker ability, the "signaling view" would suggest that hiring decisions made without this signal should leave employers disappointed. We find no statistically or economically significant treatment effects for any of these ratings. Given the 10% higher average wages in the treatment group, if employers were simply tricked into hiring worse workers generally, these higher wages should have made it even more likely to find a negative effect on ratings (Luca and Reshef, 2021). Moreover, we find that workers are hired for at least as many hours of labor, and are just as likely to be rehired.

One possible explanation for our results is that employers are simply wrong to consider resume writing quality as informative about ability. However, the "clarity view" can also rationalize our results without making this assumption.[1] Our results are consistent with jobseekers with heterogeneous productivity, where those who receive algorithmic writing assistance face a lower cost to writing clear resume which reveal their type to employers. We provide evidence for our underlying mechanism, that algorithmic writing assistance

[1]It is helpful to formalize this notion to contrast it with the more typical signaling framing of costly effort and hiring. To that end, we present a model in Appendix Section B where jobseekers have heterogeneous private information about their productivity but can reveal their type via writing a "good" resume. This model assumes that there are some workers who are unable to write a good resume, for reasons independent of their quality—e.g. due to their English language ability or lack of communication training. We show that relaxing this friction, due to the introduction of a technology which improves those workers writing, and lowers the cost of good writing can generate our findings of more hires, higher wages, and equally satisfied employers.

improves the clarity of the writing by looking at measures of writing readability (Kincaid et al., 1975). We find consistent evidence that the writing on the resumes of the treated group is easier to read than the resumes in the control group.

We perform this study in the context of a large literature on how experimentally varying applicant attributes affects callback rates (Moss-Racusin et al., 2012; Bertrand and Mullainathan, 2004; Kang et al., 2016; Farber et al., 2016). More specifically, we contribute by showing the importance of text in understanding matching (Marinescu and Wolthoff, 2020). The notion that better writing can help a reader make a better purchase decision is well-supported in the product reviews literature (Ghose and Ipeirotis, 2010) but is a relatively novel finding in labor markets.

Writing has long been used for evaluation across many spheres, for example school essays, personal statements, and cover letters in job applications. While we are not the first to investigate how writing matters to employers[2] (Sterkens et al., 2021; Martin-Lacroux and Lacroux, 2017), we believe we are the first to do so in a field experiment with natural variation in writing quality. In one related example, Sajjadiani et al. (2019) analyze resumes of applicants to public school teaching jobs and find that spelling accuracy is correlated with a higher probability of being hired. Hong, Peng, Burtch and Huang (2021) further show that workers who directly message prospective employers (politely) are more likely to get hired, but the politeness effect is muted when the workers' messages contain typographic errors. Weiss et al. (2022) conducts a lab experiment and finds that the use of AI in job-seekers' writing resulted in employer perceptions of lower competence, warmth and social desirability (however, of particular importance is that in their experiment, the use of AI was disclosed to employers).

These results come at a key time for the evolution of hiring decisions—the practical implications of these two views can inform employers who need to adapt their hiring practices to a world in which AI can provide substantial quality improvements to application materials. AI capable of generating text is already leaving its mark on labor markets (Eloundou et al., 2023; Felten et al., 2023), and understanding the role of individuals' writing abilities in predicting their quality is becoming increasingly crucial. Recent research has demonstrated that Large Language Models like ChatGPT can significantly improve worker productivity, particularly by raising bottom of the skill distribution (Noy and Zhang, 2023; Brynjolfsson et al., 2023). Similar findings have been reported in other studies on technological advancements, such as the benefit that surgical robots provide to the least proficient surgeons (Tafti, 2022). Our own findings are consistent with these results, as we observed the greatest ef-

[2] While the reason this preference exists is not known, recruiters report, anecdotally, caring about a resume's writing quality (Oreopoulos, 2011).

fects of our treatment among individuals with lower writing quality.

These results can only describe a partial equilibrium. Crowd-out effects are possible if not likely (Crépon et al., 2013; Marinescu, 2017), which are relevant to discuss the welfare implications of any market intervention. Our primary purpose is understanding how employers make decisions with respect to resumes and their role as a tool for lessening information frictions. However there are different implications to platform designers and managers if the introduction of algorithmic writing assistance increases the absolute number of matches or simply changes which jobseeker gets hired. We show that in this setting, the treatment effect is largest for jobseekers who are not competing with as many treated jobseekers, and dissipates based on how much they compete with other treated jobseekers. In the case of the clarity view, even if rolling out the algorithmic writing assistance platform-wide sees a smaller increase in matches than what is found experimentally, there are still benefits to revenue and match quality by introducing the algorithmic writing assistance as a platform-wide policy.

If the "clarity view" is more important to future hiring decisions, then any intervention that encourages better writing will be weakly beneficial for all parties. There will likely be little loss in efficiency if parties are better informed. Even better, as we show, this kind of assistance can be delivered algorithmically. These interventions are of particular interest because they have zero marginal cost (Belot et al., 2018; Briscese et al., 2022; Horton, 2017), making a positive return on investment more likely, a consideration often ignored in the literature (Card et al., 2010). On the other hand, if the "signaling view" is more important, then providing such writing assistance will mask important information and lead to poor hiring decisions, particularly if writing skills can be conceived of as social skills.[3] As for the treatment itself, unlike general advice, algorithmic interventions are adaptive. In our study, the algorithm took what the jobseeker was trying to write as input and gave targeted, specific advice that likely improved it.[4] This is likely more immediately useful than more vague recommendations, such as telling jobseekers to "omit needless words."

The rest of the paper proceeds as follows. Section 2 describes the online labor market which serves as the focal market for this experiment. Section 3 provides evidence on the relationship between writing quality and labor market outcomes from observational data from the market before any intervention. Section 4 reports the experimental results of the treatment effects on writing quality and subsequent labor market outcomes. In Section B we present a simple model that can rationalize our findings. Section 5 concludes.

[3] Deming (2017) suggests that there are labor market returns to social skills because they reduce coordination costs and are complementary to cognitive skills.

[4] The Algorithmic Writing Service does not provide whole paragraphs of text, nor is it able to be prompted.

2 Empirical context and experimental design

The setting for this experiment is a large online labor market. Although these markets are online, with a global user base, and with lower search costs (Goldfarb and Tucker, 2019), they are broadly similar to more conventional markets (Agrawal et al., 2015). Employers post job descriptions, jobseekers search among job posts and apply. Employers then decide if and who to interview or hire. Jobs can be hourly, or project based. One distinctive feature of online labor markets is that both the employer and the worker provide ratings for each other at the end of a contract.

Because of the many similarities between on and offline labor markets, a substantial body of research uses online labor markets as a setting, often for randomized experiments. Many researchers have used platforms to study the role of information in hiring, as they are difficult to study elsewhere (Stanton and Thomas, 2016; Agrawal et al., 2016; Chan and Wang, 2018; Kokkodis and Ransbotham, 2022). Online labor markets also allow researchers to broaden the range of hypotheses to test (Horton, 2010) because platforms store detailed data down to the microsecond on things like applications, text, length of time spent working on an application, speed of hire, and more.

The online labor market which serves as the setting for this experiment is a global marketplace, and not representative of, say, the US workforce. About 20% of the sample comes from anglophone countries US, Canada, UK, and Australia. However, less than 6% of the world's population comes from these Anglophone countries.[5] The sample also overweights India, which make up 17% of the global population but 24% of the workers in our sample. As a global marketplace, this market has features that distinguish it from local labor markets. All of the work is Internet-mediated, removing frictions based on geography. Still, there exist frictions based on language and communication skills, which is one of the reasons it makes a good setting to study the role of the resume in hiring. We provide summary statistics about jobs worked on the platform in Appendix Table A1. The average job lasts two months, takes 201 hours of labor with average wages of $17 per hour. Most of the work measured by the wagebill on the platform consists of hourly jobs, but fixed price jobs make up two-thirds of the total number of contracts formed.

2.1 Search and matching on the platform

A would-be employer writes job descriptions, labels the job opening with a category (e.g., "Graphic Design"), lists required skills, and then posts the job opening to the platform website. Jobseekers generally learn about job openings via electronic searches. They submit

[5] https://en.wikipedia.org/wiki/List_of_countries_and_dependencies_by_population

applications to jobs they are interested in and are required to include a wage bid and a cover letter.

In addition to jobseeker-initiated applications, employers can also use the interface to search worker profiles and invite workers to apply to particular jobs. The platform uses jobseekers' on-platform history and ratings to both recommend jobseekers directly to would-be employers and to rank them in order of relevance and quality. At no point do these algorithmic recommendations consider the writing quality of the jobseeker's resume. By using recommendation systems, algorithms can help reduce randomness in the hiring process and provide employers with quality signals about potential hires (Horton, 2017). In terms of selection, Pallais (2014) shows that employers in an online labor market care about workers' reputation and platform experience when hiring. After jobseekers submit applications, employers screen the applicants. The employers can highlight applications of interest through the platform interface to save in a separate tab, their "shortlist." Then the employer decides whether to conduct interviews, and whether to make an offer(s). If a match is formed, the platform observes the wages, hours worked, earnings, and ratings at the conclusion of the contract. Although these ratings have been shown to become inflated over time and can be distorted when they are public and reciprocal (Bolton, Greiner and Ockenfels, 2013), they are still a useful signal of worker performance (Fradkin et al., 2021; Cai et al., 2014). We consider the impact of the treatment to the public and private numerical ratings the employers give to the workers, as well as the "sentiment" of the written text of reviews, which are less prone to inflation (Filippas et al., 2022).

2.2 Experimental intervention at the resume-writing stage of profile creation

When new jobseekers sign up to work on the platform, their first step is to register and create their profile. This profile serves as the resume with which they apply for jobs. This profile includes a list of skills, education, and work experience outside of the platform. It also includes a classification of their primary job category (e.g., "Graphic Design"), mirroring what employers select when posting a job. The interface consists of a text box for a profile title and a longer text box for a profile description. Their finished profile will include their profile description and a "profile hourly wage," which is the wage they offer to employers searching for workers.

During the experimental period, jobseekers registering for the platform were randomly assigned to an experimental cell. The experimental sample comprises jobseekers who joined the platform between June 8th and July 14th, 2021. For treated jobseekers, the text boxes for the profile description are checked by the Algorithmic Writing Service. Control jobseekers received the status quo experience. The experiment included 480,948 jobseekers, with 50% allocated to the treated cell. Table 1 shows that it was well-balanced and the balance of pre-treatment covariates was consistent with a random process.[6]

2.3 The algorithmic writing assistance

Words and phrases that a language model determines to be errors are underlined by the Algorithmic Writing Service. See Figure 1 for an example of the interface. By hovering a mouse cursor over the underlined word or phrase, the user sees suggestions for fixing spelling and grammar errors. The Algorithmic Writing Service also advises on punctuation, word choice, phrase over-use, and other attributes related to clarity, engagement, tone, and style.

According to the Algorithmic Writing Services website, the software uses a combination of transformer models and rule-based systems to provide its recommendations. Unlike Large Language Models like ChatGPT or BingChat, this system is not generative—it cannot be prompted or asked questions, it simply takes the text the user has provided and suggests improvements to it.

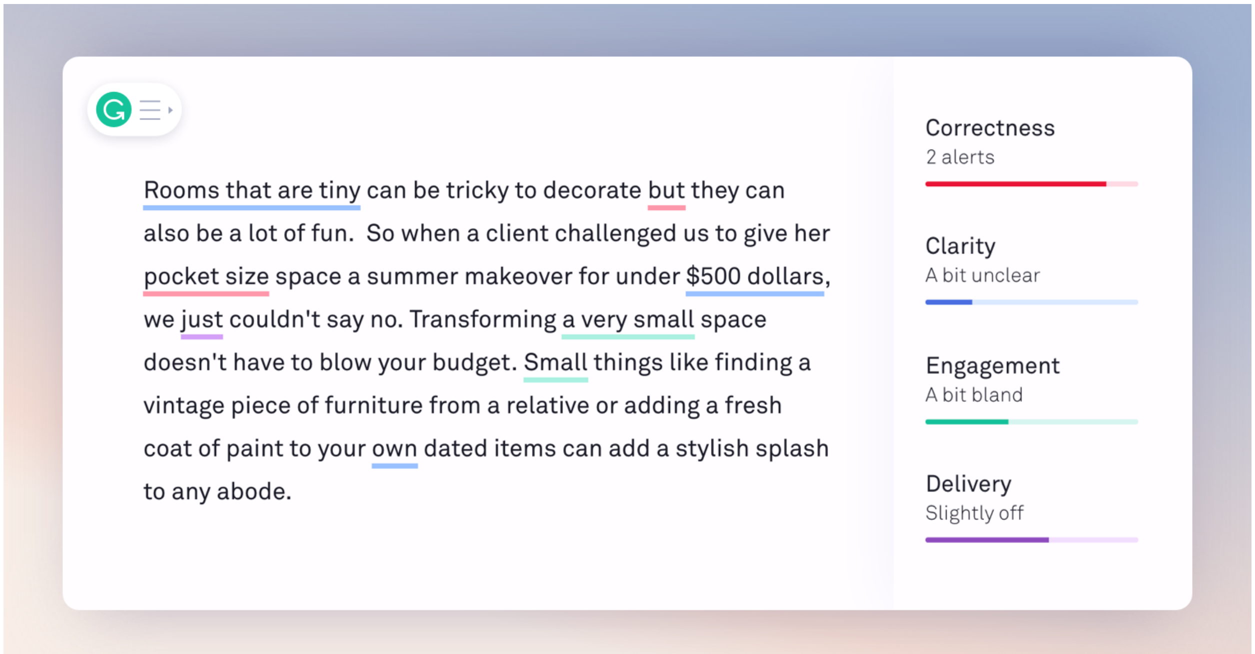
2.4 Platform profile approval

Of the experimental sample, 46% of workers allocated into the experiment upon registration complete and submit their profiles. When jobseekers finish setting up their profiles, they have to wait to be approved by the platform. The platform approves jobseekers who have filled out all the necessary information, uploaded ID, and provide bank details so they can be paid. The platform can reject jobseekers at their discretion. However, platform rejection is somewhat rare. About 10% of profiles are rejected, usually as a part of fraud detection or because the jobseekers leave a completely empty profile. About 41% of workers who begin registering get all the way through the approval process.

As approval is downstream of profile creation, this step creates a potential problem for interpreting any intervention that changes profile creation. For example, it could be that the treatment leads to a greater probability of platform approval. Or, the treatment could have made jobseekers more likely to complete the registration process and submit their profile, both of which could effect hiring. While unlikely given the mechanistic rules the platform applies, this is possible, and we investigate this potential issue in multiple ways.

[6] In Appendix Figure A1 we show the allocations by treatment status over time and find they track closely.

Figure 1: Example of the Algorithmic Writing Service's interface showing suggestions on how to improve writing



Notes: Example of the Algorithmic Writing Service applied to a paragraph of text. To receive the suggestions, users hover their mouse over the underlined word or phrase. For example, if you hover over the first clause "Rooms that are tiny" underlined in blue, "Tiny rooms" will pop up as a suggestion.

First, we check whether there is any evidence of selection and find no evidence that treated jobseekers are no more likely either to submit their profiles and or to receive approval.[7]

Second, in our main analysis, we condition on profile approval in our regressions. We also perform robustness checks where we report the same analysis not conditioned on profile approval and where we control for profile approval as a covariate. All findings are robust to these strategies, a result described in Section 4.5.

2.5 Description of data used in the analysis

The dataset we use in the analysis consists of the text of jobseekers' resumes as well as all of their behavior on the platform between the time they registered—between June 8th and July 14th 2021—and August 14th, 2021. We construct jobseeker level data, including the title and text of their profile, the number of applications they send in their first month on the platform, the number of invitations they receive to apply for jobs, the number of interviews they give, and the number of contracts they form with employers. The most common

[7]See Appendix Table A3 for regression output.

Table 1: Comparison of jobseeker covariates, by treatment assignment

	Treatment mean: $\bar{X}_{TRT}$	Control mean: $\bar{X}_{CTL}$	Difference in means: $\bar{X}_{TRT} - \bar{X}_{CTL}$	p-value
Full sample description: N = 480,948				
Resume submitted	0.456 (0.001)	0.455 (0.001)	0.001 (0.001)	0.450
Platform approved	0.407 (0.001)	0.406 (0.001)	0.002 (0.001)	0.185
Resume length	32.911 (0.116)	32.860 (0.117)	0.051 (0.165)	0.756
Profile hourly rate	18.843 (0.126)	18.917 (0.126)	-0.075 (0.178)	0.676
Flow from initial allocation into analysis sample				
	Treatment (N)	*Control (N)*	*Total (N)*	
Total jobseekers allocated	240,231	240,717	480,948	
↪ who submitted their profiles	109,638	109,604	219,242	
↪ and were approved by the platform	97,859	97,610	195,469	
↪ with non-empty resumes	97,479	97,221	194,700	
Pre-allocation attributes of the analysis sample: N = 194,700				
From English-speaking country	0.182 (0.001)	0.183 (0.001)	-0.002 (0.002)	0.361
US-based	0.141 (0.001)	0.143 (0.001)	-0.002 (0.002)	0.221
Specializing in writing	0.173 (0.001)	0.176 (0.001)	-0.003 (0.002)	0.105
Specializing in software	0.115 (0.001)	0.115 (0.001)	0.000 (0.001)	0.771
Resume length	70.393 (0.222)	70.260 (0.222)	0.134 (0.314)	0.670

Notes: This table reports means and standard errors of various pre-treatment covariates for the treatment group and the control group. The first panel shows the post-allocation outcomes of the full experimental sample i) profile submission, ii) platform approval, iii) length of resume in the number of words, iv) profile hourly wage rate in USD. The means of profile hourly rate in treatment and control groups are only for those profiles which report one. The reported p-values are for two-sided t-tests of the null hypothesis of no difference in means across groups. The second panel describes the flow of the sample from the allocation to the sample we use for our experimental analysis. The complete allocated sample is described in the first line, with each following line defined cumulatively. The third panel looks at pre-allocation characteristics of the jobseekers in the sample we use for our analysis, allocated jobseekers with non-empty resumes approved by the platform. We report the fraction of jobseekers i) from the US, UK, Canada, or Australia, ii) from the US only, iii) specializing in writing jobs, iv) specializing in software jobs, and v) the mean length of their resumes in the number of words.

categories listed as worker's primary job categories are, in order of frequency, Design & Creative, Writing, Administrative Support, and Software Development.

In Table 1 we present summary statistics about the jobseekers in the full experimental sample, as well as the sample conditioned on platform approval. Jobseekers with writing as their primary area of work make up 17% of the sample. Only 14% of jobseekers are based in the US, and over 80% are based in a country where English is not the native language.

2.6 Constructing measures of writing quality

We do not observe the changes that the Algorithmic Writing Service suggested—we simply observe the resumes that result. As such, we need to construct our own measures of writing quality to determine if the treatment was delivered.

Algorithmic Writing Service provides text improvement suggestions along several dimensions. We measure writing quality of each resume by using a different service, LanguageTool, an open-source software that uses language models to determine various types of writing errors.[8] LanguageTool is a rule-based dependency parser that identifies errors (rule violations) and categorizes them. Some example categories include "Nonstandard Phrases," "Commonly Confused Words," "Capitalization," and "Typography." For example, the non-standard phrase "I never have been" would be flagged with a suggestion to replace it with "I have never been."[9] Our primary measures of writing quality are the error rates for each of these error types, as well as the overall error rate. The error rate is determined by totaling the number of all error types classified by LanguageTool, normalized by number of words in the resume.

3 Observational results

Before presenting results of the field experiment, we explore the relationship between resume writing quality and hiring using observational data from this market.

3.1 The association between writing quality and hiring probabilities

More writing errors are associated with lower hiring probabilities in the observational data. In Figure 2, we plot jobseekers' hiring outcomes versus the error rate, controlling for the length of the resume. The sample is the resumes of all jobseekers who registered for the platform over the month of May 7th through June 7th, 2021, prior to the experiment. The distribution of the error rate is very right skewed—over 95% of jobseekers' resumes have error rates of less than 25%. In Figure 2, the x-axis is the deciles of error rate, truncated to include only jobseekers whose resumes have error rates of less than 25%. The y-axis is the

[8]This is a different software than the Algorithmic Writing Service.

[9]For a more detailed explanation of all of the rule categories, see Appendix Table A4.

Figure 2: Association between resume error rate and if a jobseeker is hired in observational data

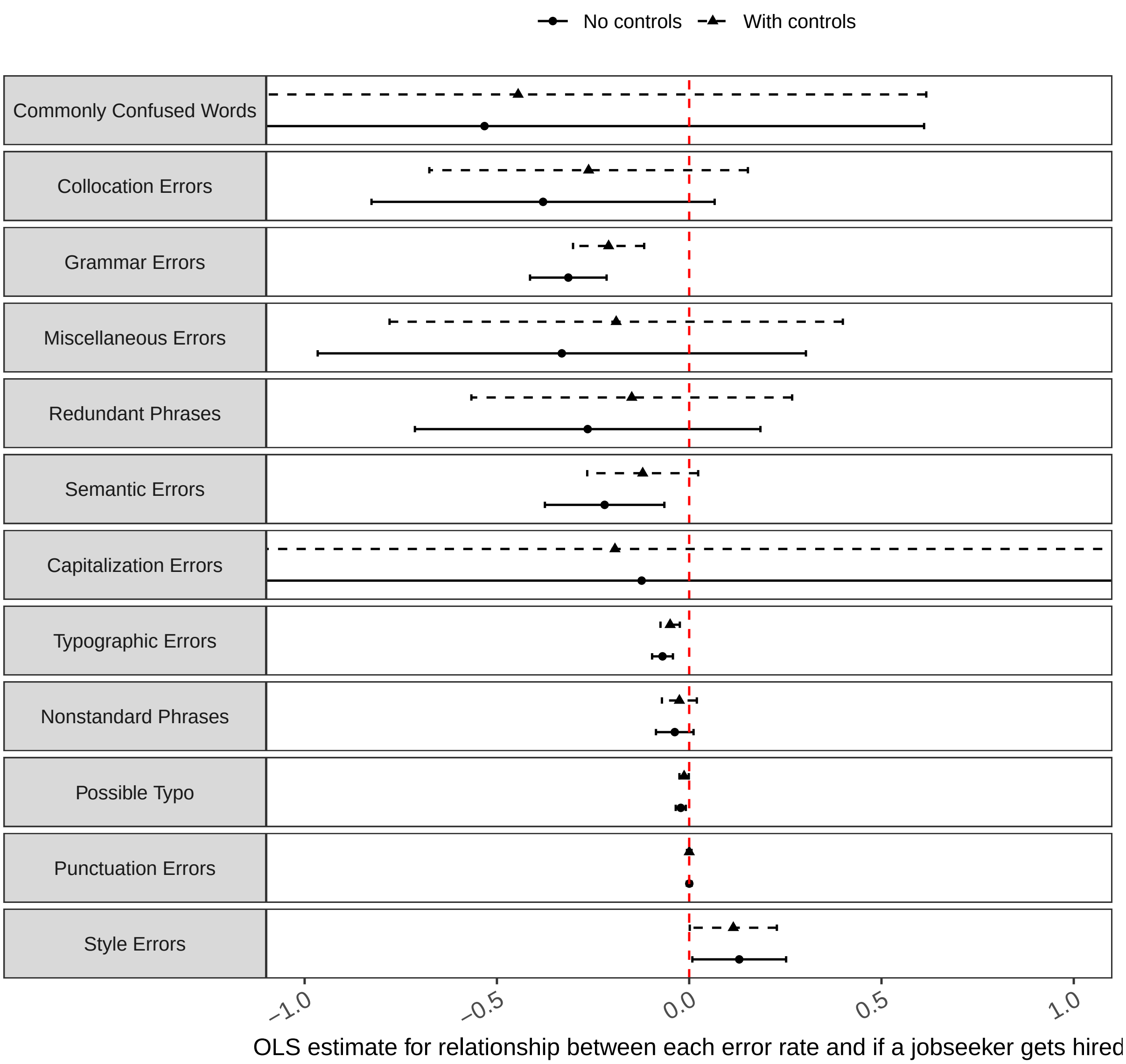
Notes: These data show the relationship between the error rate on a jobseekers' resume with the probability they were hired within a month, controlling for resume length. The error rate is determined by totaling the number of all error types classified by LanguageTool, normalized by number of words in the resume. A 95% confidence interval is plotted around each estimate. The sample is all new jobseekers who were approved by the platform between June 1st and June 7th, 2021, with resumes of more than 10 words. The x-axis is error rate deciles on the sample of resumes where the error rate is less than 25%.

residuals from regressing the error rate on whether or not the jobseeker is hired, controlling for number of words in the resume. Generally, jobseekers with resumes with a lower error rate (deciles to the left of the plot) are more likely to be hired.

In order to unpack the various types of errors, in Figure 3 we show the correlation between hiring outcomes and each individual type of language error in the observational data.[10] In the first specification, we show the correlation between the error rate for the various types of language errors and an indicator for whether or not the jobseeker is ever hired in their first 28 days after registering for the platform . In the second specification, the outcome is simply the number of contracts formed over the jobseeker's first month. In the second specification, we control for the jobseekers' profile hourly rate and primary category of work.

[10]In Appendix Table A5 we show the table of these estimates. In Appendix Table A2 we summarize the frequency of these error types in the observational data.

Figure 3: Relationship between writing error rate and if a jobseeker is hired in observational data



Notes: These data show the relationship between the error rate on a jobseekers' resume with the probability they were hired within a month. Specification with controls include resume length, jobseeker category, and profile hourly rate. The error rate is determined by the number of each error type classified by LanguageTool, normalized by number of words in the resume. Error type definitions can be found in Appendix Table A4. A 95% confidence interval is plotted around each estimate. The sample is all new jobseekers who were approved by the platform between June 1st and June 7th, 2021. Regression tables the plot is based on can be found in Appendix Table A5.

Resumes with more per word grammar errors, typos, typography errors, and miscellaneous errors are all hired less. This linear model places some unreasonable assumptions like constant marginal effects on the relationship between various writing errors and hiring. There may be interactions between these error types. However, it is still useful to summarize the relationships. We can see generally negative relationships between writing error rate and hiring. In the second specification where we add controls, we see coefficients get smaller in magnitude as we would expect, but the significance does not disappear. For robustness we repeat these analysis in levels in Appendix Table A7.

In terms of magnitude, one additional error of any type is associated with that jobseeker being hired 1.4% less. In Appendix Table A6 we show the relationship between total number of errors and hiring outcomes and report these results in both levels and normalized by resume length. The negative relationship between writing errors and hiring persists in all specifications.

4 Experimental results

We look at three main kinds of experimental results. First, we examine how the treatment affected the text of resumes. Next, we look at employment outcomes for those treated workers. Third, we will look at how the treatment impacted the quality of work, as assessed by employer reviews and whether or not a worker is rehired.

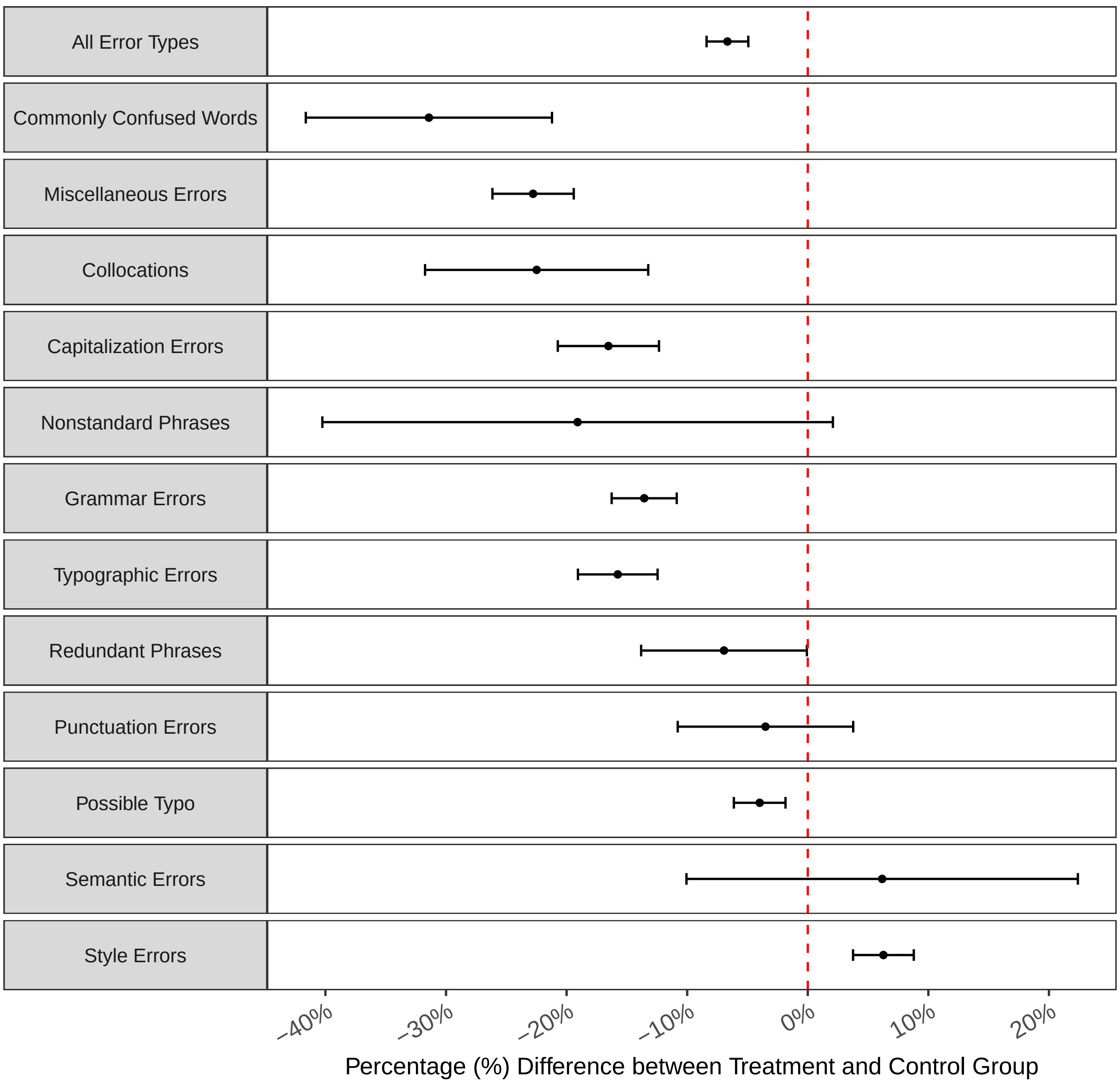
4.1 Algorithmic writing assistance improved writing quality

The first step of our analysis is to measure the effect that the Algorithmic Writing Service has on the writing of the resumes in the treatment group. We begin by analyzing the effect of the treatment on all types of writing error rates, as defined by LanguageTool. Figure 4 displays the effect of treatment on the number of each type of writing error, normalized by resume length.[11] For treatment effects measured in percentage terms, we calculate the standard errors using the delta method.

In the first facet of Figure 4, we find that jobseekers in the treatment group made 5% fewer errors in their resumes. Breaking these down by error type, we find that jobseekers in the treatment group had a significantly lower rate of errors of the following types: capitalization, collocations, commonly confused words, grammar, possible typos, miscellaneous, and typography. We find larger treatment effects for errors associated with writing clarity than for many others. For example, two of the largest magnitudes of differences in error rate

[11] The treatment had no effect on the length of resumes—see Appendix Table A8.

Figure 4: Effect of the algorithmic writing assistance on resume error rates



Notes: This plot shows the effect of the treatment on various writing errors in jobseekers' resumes, normalized by number of words in the resume. Point estimates are the percentage change in the dependent variable versus the control group for the treatment groups. A 95% confidence interval based on standard errors calculated using the delta method is plotted around each estimate. The experimental sample is of all new jobseekers who registered and were approved for the platform between June 8th and July 14th, 2021, and had non-empty resumes, with N = 194,700. Regression details can be found in Appendix Tables A9 and A10. Bonferroni Corrected standard errors can be found in Appendix Table A11.

were commonly confused words and collocations, where two English words are put together that are not normally found together. Interestingly, the treatment group had more "style" errors, paralleling our results from the observational data (see Table A5).

4.1.1 Heterogeneous treatment effects to writing quality

A natural question is which jobseekers benefited most from the treatment. Appendix Table A14 interacts pre-randomization jobseeker attributes with the treatment. We can see that jobseekers from the US or from English-speaking countries,[12] all have fewer errors in "levels."

The treatment negatively impacted the writing error rate of all subgroups by country of origin. We find that jobseekers from non-native English-speaking countries experience significantly larger treatment effects to their error rate. Still, effects to their Anglophone counterparts are negative and significant.

In Appendix Table A15 we focus on jobseekers who list their primary category of work as "Writing" and in Column (1) show that the treatment even has a significant impact on the writing on writers' resumes.

4.2 Effects to workers employment outcomes

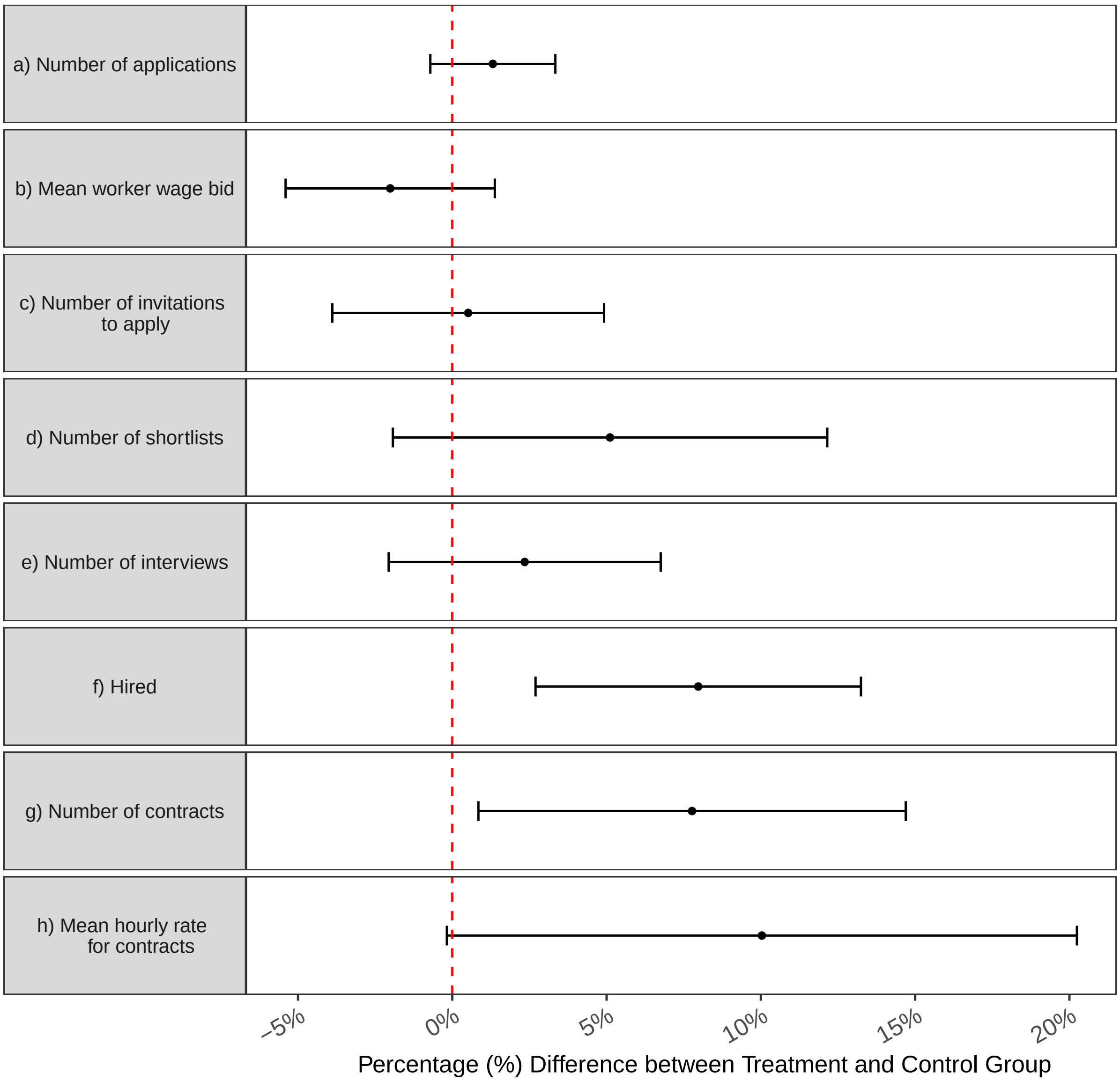
Access to the treatment impacted whether or not jobseekers were hired. Figure 5 summarizes the treatment effects on the primary hiring outcomes.

4.2.1 Treated workers did not change their job search strategy or behavior

The potential for the treatment to impact jobseeker search behavior or intensity could complicate our desire to focus on employer decision-making. Job applications have been shown to be costly (Abebe, Caria and Ortiz-Ospina, 2021) and job search intensity could depend on jobseekers expectation of their own hireability. It is possible that treated jobseekers realized they were in an experiment and increased their search efforts, knowing they had higher quality resumes. In that case, we could not interpret our treatment effect as being driven by employers' improved perceptions of treated jobseekers. We therefore plot the percentage change in job search metrics for jobseekers in the treatment versus those in the control group in Figure 5a) and Figure 5b) and find no evidence that jobseekers changed their search behavior.

[12] We define whether a jobseeker is from an Anglophone country, by whether they login to the platform from USA, UK, Canada, Australia, or New Zealand.

Figure 5: Effect of algorithmic writing assistance on hiring outcomes



Notes: This analysis looks at the effect of treatment on hiring outcomes on jobseekers in the experimental sample. The x-axis is the difference in the mean outcome between jobseekers in the treated group and the control group. A 95% confidence interval based on standard errors calculated using the delta method is plotted around each estimate. The experimental sample is of all new jobseekers who registered and were approved for the platform between June 8th and July 14th, 2021, and had non-empty resumes, with N = 194,700. Regression details on the number of applications and wage bid can be found in Table 2. Regression details on invitations, interviews, hires, and the number of contracts can be found in Table 3. Regression details on hourly wages can be found in Table 4.

Table 2 provides regression results for these effects of the treatment on jobseekers' search behavior. In Column (1) the outcome is the number of applications a jobseeker sends out over their first 28 days after registering. In the control group, jobseekers send on average 2 applications in their first month on the platform. We find no effect of the treatment on the total number of applications sent.

In Table 2 Column (2), the outcome is the mean wage bid proposed by the jobseekers on those applications. We find that treated jobseekers do not apply to more hourly jobs than those in the control group. They also could have bid for higher wages knowing they had better-looking resumes. In Table 2 Column (3), the outcome is the mean wage bid proposed by the jobseekers on those applications. Average wage bids in both the treatment and control groups were $24 per hour. This lack of impact to jobseeker's behavior makes sense as jobseekers were not made aware of the fact that they were in an experiment.

Table 2: Effects of writing assistance on jobseekers' application behavior

	Dependent variable:		
	Num Applications	Num Hourly Applications	Mean Hourly Wage Bid
	(1)	(2)	(3)
Algo Writing Treatment	0.023	0.012	-0.492
	(0.018)	(0.011)	(0.427)
Constant	1.768***	0.919***	24.425***
	(0.013)	(0.008)	(0.302)
Observations	194,700	194,700	59,854
R^2	0.00001	0.00001	0.00002

Notes: This table analyzes the effect of the treatment on jobseekers' application behavior. The experimental sample is made up of all new jobseekers who registered and were approved by the platform between June 8th and July 14th, 2021 and had non-empty resumes. The outcome in Column (1) is the number of total applications a jobseeker sent out between the time the experiment began and one month after it ended. The outcome in Column (2) is the number of specifically hourly applications sent out in that same time period. The outcome in Column (3) is the mean hourly wage bid they proposed for those hourly jobs, and the sample narrows to only jobseeker who submitted at least one application to an hourly job.
Significance indicators: $p \leq 0.10 : *$, $p \leq 0.05 : **$ and $p \leq .01 : ***$.

4.2.2 The treatment did not affect employer recruiting

Employers are able to seek out workers using the platform's search feature to invite jobseekers to apply to their job openings. In Figure 5c), the outcome is the number of invitations to apply for a job that the jobseeker receives in their first month. We find no effect of the treatment on employer invitations.

This result makes sense given that our experimental sample consists of only new jobseekers to the platform. New entrants almost never appear in the search results when employers search for jobseekers, given that their rank is determined by their platform history. Given that the search feature is how employers find jobseekers to invite to jobs, we would not expect the treatment to affect invitations to apply. Table 3 Column (1) provides the details of this regression.

After jobseekers apply, employers can sort through the applications to their job and highlight applications they are especially interested in through a feature called shortlisting. In Figure 5d) we observe that jobseekers in the treatment group had applications shortlisted 5% more than jobseekers in the control group, although this effect is not significant. Table 3 Column (2) provides the details of this regression.

The treatment had no significant impact to number of interviews. In Figure 5e), we show the effect of the treatment on number of interviews. Interviews, while technically feasible, are rare on this platform, and do not correspond the types of interviews given in offline labor markets. Here, an interview is defined as any correspondence via message between the employer and applicant, prior to an offer being made. In the control group the average jobseeker gives 0.18 interviews over the course of their first month after registering, with the treatment group receiving 2.5% more interviews. Table 3 Column (3) provides the details of this regression.

4.2.3 Treated jobseekers were more likely to be hired

The treatment raised jobseekers' hiring probability and the number of contracts they formed on the platform. In Figure 5f), the outcome is a binary indicator for whether or not a jobseeker is ever hired in their first 28 days on the platform. During the experiment, 3% of jobseekers in the control group worked at least one job on the platform. Treated jobseekers see an 8% increase in their likelihood of being hired in their first month on the platform.

Jobseekers in the treated group formed 7.8% more contracts overall. In Figure 5g), the outcome is the number of contracts a jobseeker worked on over their first month. In Table 3 Column's (4) and (5) we report these results in levels.

4.2.4 Hourly wages in formed matches were higher

Treated workers had 10% higher hourly wages than workers in the control group. In Figure 5h), the outcome is the mean hourly rate workers earned in jobs they worked over their first month on the platform.[13]

[13] Hourly wage rates for new entrants are not representative of rates on the platform. If a new entrant gets hired for their first job, they tend to experience rapid wage growth.

Table 3: Effect of algorithmic writing assistance on hiring outcomes

	Dependent variable:				
	Num Invitations	Num Shortlists	Num Interviews	Hired x 100	Num Contracts
	(1)	(2)	(3)	(4)	(5)
Algo Writing Treatment	0.001	0.002	0.004	0.247***	0.004**
	(0.003)	(0.001)	(0.004)	(0.080)	(0.002)
Constant	0.142***	0.039***	0.178***	3.093***	0.047***
	(0.002)	(0.001)	(0.003)	(0.057)	(0.001)
Observations	194,700	194,700	194,700	194,700	194,700
R^2	0.00000	0.00001	0.00001	0.00005	0.00003

Notes: This analysis looks at the effect of treatment on hiring outcomes on jobseekers in the experimental sample. The Column (1) outcome Invitations is the number of times they were recruited to a job over their first month. Column (2) is the number of times their application was shortlisted over that month. Column (3) is the number of interviews they gave over that month. Column (4) defines Hired x 100 as one hundred times the probability the jobseeker was hired over that month. Column (5) defines Number of Contracts as the number of unique jobs they work over the month after they register for the platform. The experimental sample is of all new jobseekers who registered and were approved for the platform between June 8th and July 14th, 2021 and had non-empty resumes. Significance indicators: $p \leq 0.10 : *$, $p \leq 0.05 : **$ and $p \leq .01 : ***$.

In Table 4 Column (1) we show that in the control group, workers on average made \$17.25 per hour. In the treatment group, workers made \$19.01 per hour, with a p-value of 0.05. Since workers did not bid any higher, it is possible that employers are hiring more productive workers, or that they thought the treated workers were more productive. If that is the case, the “signaling view” would predict that employers would then be disappointed with the workers they hired, which we should be able to observe in worker ratings.

Because these effects are downstream of hiring, these higher wages could be a result of bargaining or due to a composition effect. We find that the initial wage bids are almost always the same as the hourly wage and there is very little evidence of bargaining. In this sample of hires, in only 0.2% of contracts the freelancer proposes more than one bid before being hired. Initial wages and bids are 92% correlated for hourly jobs and 95% correlated for fixed price jobs. In Table 4 Column (2) we regress the treatment on an indicator variable defined as 1 if the jobseekers’ initial wage bid is equal to the hourly wage they are hired for, and 0 if not. Using this definition as well, we see no evidence that the treatment increased bargaining.

Taken together with the fact that there is no effect of the treatment to asking wage bids, as we show in Table 2, this evidence points to the increase in hourly wages being driven by a composition effect.

Table 4: Effect of algorithmic writing assistance on average contract wages

	Dependent variable:	
	Hourly wage rate	I(Bargaining)
	(1)	(2)
Algo Writing Treatment	1.763**	−0.027
	(0.834)	(0.020)
Constant	17.247***	0.277***
	(0.611)	(0.015)
Clustered SEs	X	X
Observations	3,305	1,949
R^2	0.001	0.001

Notes: This analysis looks at the effect of treatment on hourly wages of contracts for jobseekers in the experimental sample, conditional on a hire. The sample is at the job level, and we cluster standard errors at the worker level. The outcome in Column (1), hourly wage rate, is defined as the max hourly wage rate a worker recieves for that job. In Column (2) the outcome is an indicator which is 1 if the jobseeker's wage bid is not equal to the wage they are hired at, and 0 if else. The experimental sample is of all new jobseekers who registered and were approved for the platform between June 8th and July 14th, 2021 and had non-empty resumes, for all jobs they worked within 28 days of registering for the platform. Significance indicators: $p \leq 0.10 : *$, $p \leq 0.05 : **$ and $p \leq .01 : ***$.

Table 5: Effects of algorithmic writing assistance on hours worked and rehires

	Dependent variable:	
	Hours worked	Ever rehired
	(1)	(2)
Algo Writing Treatment	0.412	−0.003
	(0.303)	(0.007)
Constant	2.649***	0.079***
	(0.214)	(0.005)
Observations	194,700	6,263
R^2	0.00001	0.00003

Notes: This table analyzes the effect of the treatment on measures of hours worked and rehires. In Column (1) the outcome is the number of total hours worked by a worker in their first 28 days on the platform. In Column (2) the outcome is the fraction of workers who are ever rehired for different jobs by the same employer, conditional on jobseekers working at least one job. The experimental sample is of all new jobseekers who registered and were approved by the platform between June 8th and July 14th, 2021 and had non-empty resumes. Significance indicators: $p \leq 0.10 : *$, $p \leq 0.05 : **$ and $p \leq .01 : ***$.

4.2.5 Hours worked were unaffected by the treatment

After examining the effects of the treatment on hiring outcomes, we now turn our attention to employer satisfaction with the workers' labor. One proxy for employer satisfaction is each worker's total number of hours worked, as this can be an indication of how much demand there is for their services. In Table 5 Column (1) we show that treated workers worked no fewer hours than workers in the control group. This sample for this analysis is the entire experimental sample who finished registration and were approved by the platform. The average worker in the control group only works for 2.6 hours during their first month on the platform. However, among those who are ever hired, the average worker in the control group works 238 hours.

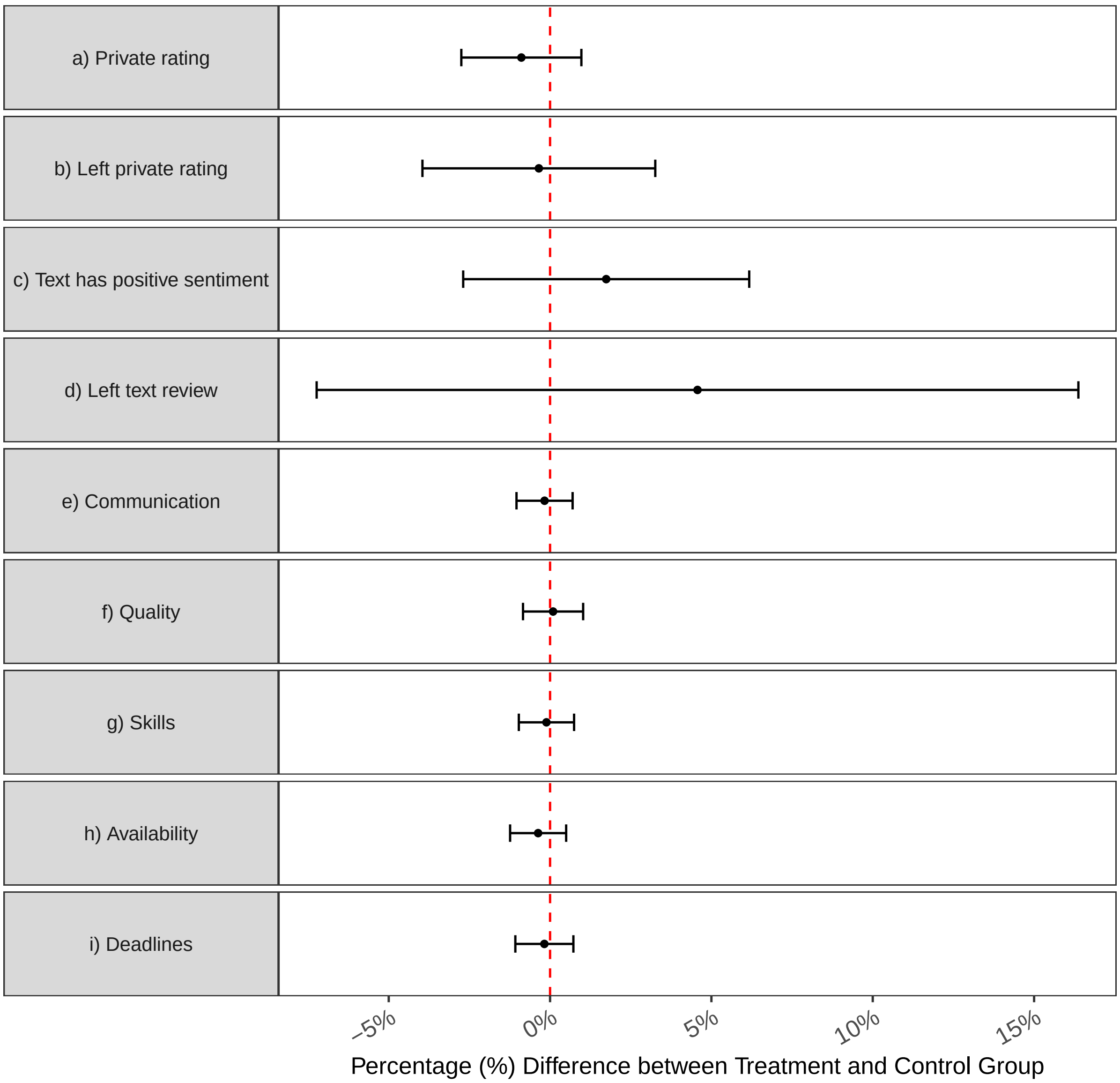
Lastly in Column (2) we show the impact of the treatment to the fraction of workers that are ever rehired. Unlike the other outcomes, rehires are conditional on a worker being hired at least once over their first month on the platform. All jobseekers in this sample have been hired at least once, and the outcome "ever rehired" is 1 if the jobseeker is ever hired a second time by their first employer and 0 if they are only hired once. About 8% of all workers are rehired by the same employer at least once over the course of the experiment. This fraction does not differ in the treatment and control group.

4.3 Employers satisfaction was unaffected by the treatment

At the end of every contract, employers rate and review the workers by reporting both public and private rating to the platform. Private ratings are not shared with the worker. In the control group, workers had an average private rating of 8.63. In Figure 6a) we show that treated workers who formed any contracts over the experimental period did not have statistically different private ratings than workers in the control group. In Column (1) of Table 6 report the results from this regression. We show that workers in the treated group have an average private rating of 8.56 with a standard error of 0.08. We may also worry that if employers are less happy with the workers quality or productivity, that they may be more or less likely to leave a review at all. Figure 6b) we show that workers in the treatment group are not more or less likely to receive any rating than workers in the control group.

When the employers give these ratings they are also able to leave text reviews. While numerical ratings have become inflated in recent years, Filippas et al. (2022) show that the sentiments associated with the text of reviews has increased significantly less over time. This means that text reviews are likely more informative about the workers' quality than the numerical ratings. We use a BERT text classification model (HF canonical model maintainers, 2022) to label each review as having positive or negative sentiment. These clas-

Figure 6: Effect of algorithmic writing assistance on ratings



Notes: This analysis looks at the effect of treatment on ratings outcomes on jobseekers in the experimental sample. Private ratings are on a scale from one to ten. Communication, Quality, Skills, Availability, and Deadlines ratings are public and left as star ratings, on a scale from one to five. The x-axis is the difference in the mean outcome between jobseekers in the treated group and the control group. A 95% confidence interval based on standard errors calculated using the delta method is plotted around each estimate. The experimental sample is of all new jobseekers who registered and were approved for the platform between June 8th and July 14th, 2021, and had non-empty resumes that were hired in their first month on the platform, with $N = 4,250$. Regression details on private ratings and text reviews can be found in Table 6. Regression details on public ratings can be found in Appendix Table A18.

Table 6: Effect of algorithmic writing assistance on contract ratings

	Dependent variable:			
	Private rating	Positive text review	Left any rating	Left any text review
	(1)	(2)	(3)	(4)
Algo Writing Treatment	−0.077	0.015	−0.002	0.006
	(0.082)	(0.019)	(0.012)	(0.008)
Constant	8.633***	0.859***	0.624***	0.138***
	(0.059)	(0.014)	(0.008)	(0.006)
Observations	4,250	1,185	6,263	6,263
R^2	0.0002	0.001	0.00001	0.0001

Notes: This analysis looks at the effect of treatment on contract outcomes for jobseekers in the experimental sample. Column (1) defines private rating as the mean private rating on all jobs given by employers to the workers after the job ended, at the worker level. In Column (2) we take the text of the reviews left by employers on each job and use sentiment analysis (model: distilbert-base-uncased-finetuned-sst-2-english) to impute whether the review is positive, neutral, or negative, labeled one if it is positive or neutral. The outcome is the mean of these ratings over all contracts in the sample. Column (3) is the percentage of contracts worked where the freelancer recieved any private rating. And Column (4) is the percentage of contracts worked where the freelancer recieved any text based review. The experimental sample is of all new jobseekers who registered and were approved for the platform between June 8th and July 14th, 2021 and had non-empty resumes, for all jobs they worked within 28 days of registering for the platform. Significance indicators: $p \le 0.10 : *$, $p \le 0.05 : **$ and $p \le .01 : ***$.

sifications are significantly correlated with the private ratings, with a Pearson correlation coefficient of 0.54. In Figure 6c) we show that the treated workers' average text reviews are not statistically different from the average sentiment of the reviews for control workers. In Figure 6d) we show that workers in the treatment group are not more or less likely to receive any text review than workers in the control group. Results from these regressions can be found in Table 6.

In Figure 6e) through i) we report the results of the effect of the treatment on the employers' public ratings of the workers. Each outcome is a public rating the employers give to the workers at the end of a contract. Employers rate the workers communication, skills, quality of work, availability, cooperation, and ability to make deadlines. Each rating is given on a five point scale. There is less variation in the public ratings than in the private ones, and the average rating for each attribute is over 4.75 stars. Like the private ratings, there are no significant effects of the treatment to any of the ratings, including to workers' communication skills. And the point estimate of the treatment effect to the quality of the work done is even positive. Results from these regressions can be found in Appendix Table A18.

4.3.1 How much power do we have to detect worse contractual outcomes?

Given the null effect of the treatment to ratings, a natural question is how much power is available to detect effects. While we do find a substantial increase in hiring—8%—these marginal hires are mixed in with a much larger pool of "inframarginal" hires that would likely be hired anyway.How much worse could those marginal applicants have been and still get our results to private ratings in the treatment?

Let I indicate "inframarginal" jobseekers who would have been hired in the treatment or control. Let M indicate "marginal" jobseekers who are only hired in the treatment. For workers in the control group, the average private rating will be $\bar{r}_C = \bar{r}_I$. But for the treatment, the mean rating is a mixture of the ratings for the inframarginal and the ratings for the induced, marginal applicants, and so

$$\bar{r}_T = \frac{\bar{r}_I + \tau \bar{r}_M}{1+\tau} \tag{1}$$

where τ is the treatment effect. We assume no substitution, making our estimates conservative. The sampling distribution of the mean rating for the marginal group is

$$\bar{r}_M = \frac{\bar{r}_T(1+\tau) - \bar{r}_C}{\tau} \tag{2}$$

Our course, $\bar{r}_T$, τ and $\bar{r}_C$ are all themselves random variables. Furthermore, they are not necessarily independent. To compute the sampling distribution of $\bar{r}_M$, we bootstrap sample both the hiring regressions and the private feedback regressions on the experimental sample.[14] Because we do not have feedback on workers who are never hired, we use the estimates values to calculate $\bar{r}_M$. Figure 7 shows the sampling distribution of $\bar{r}_M$.

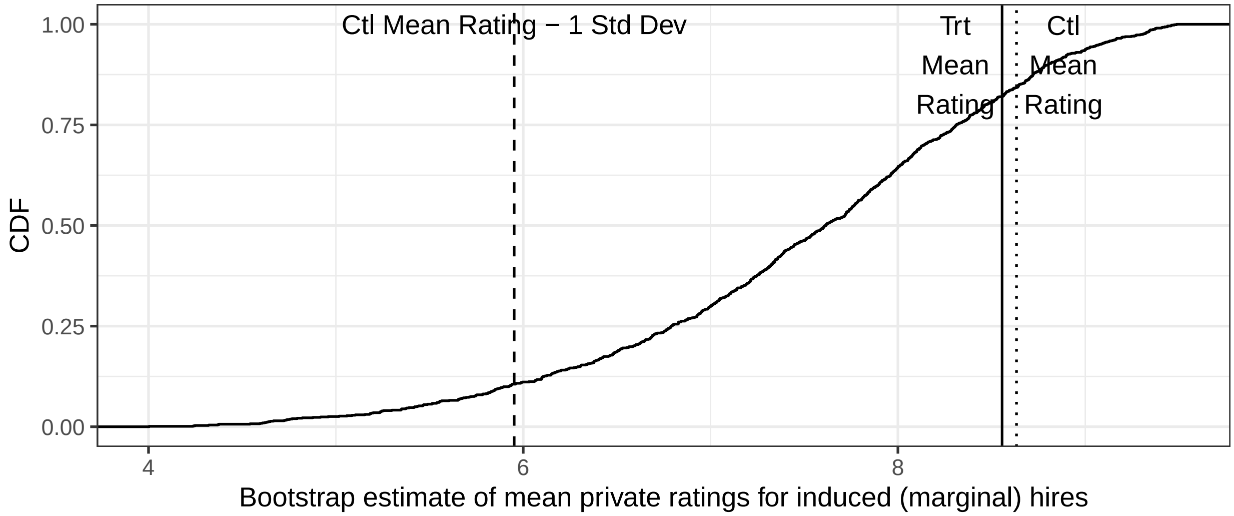
The treatment actual rating is plotted as a dotted line and control actual rating is plotted as a solid vertical line. The distribution is centered at these mean values.

The dashed line indicates the control mean rating minus one standard deviation in the private ratings (where the standard deviation is 2.4). Comparing this value to the distribution of $\bar{r}_M$, this value (at the dashed line) lies at less than 0.1 of the density. In short, it would be quite surprising for us to get the results we have—an 8% increase in hires and no different ratings if the actual marginal hires were a standard deviation worse.

Due to concerns about the loss of information in ratings caused by ratings inflation, it is reasonable to question the level of variation that could realistically be observed, even

[14]We define this sample as the workers allocated into the experiment who were approved by the platform and had non-empty resumes. From this we bootstrap sample with replacement. We run the hiring regressions on this sample and the ratings regressions on the same samples, narrowed to only those workers who were ever hired.

Figure 7: Sampling distribution of the private ratings of marginal hired jobseekers



in the presence of real effects. We do find variation in the ratings given to workers on the platform. In particular, workers with profiles written in a language other than English have an average private rating of 7.9 out of 10, which is lower than the average rating of 8.6 out of 10 for workers with profiles in English. Among workers with profiles in English, those based in the US have an average rating of 9.08 (with a standard deviation of 2.8), while workers from outside the US have an average rating of 8.46 (with a standard deviation of 2.14).

We can conduct a power analysis to determine the smallest effect size we could rule out with confidence. With 80% power and a 0.05 significance level, we can rule out any effects larger than 0.2 of a standard deviation. The overall standard deviation of ratings is 2.4, so an effect size of 0.2 standard deviations corresponds to a difference of 0.48 in ratings. This effect size is within the range of variation in ratings that we see within the data. Therefore, we can be reasonably confident that our study design would have been able to detect effects of practical significance.

4.4 Heterogeneous treatment effects to hiring and ratings

We have already shown above in Appendix Table A14 that the treatment disproportionately impacted the error rate in non-native English speakers' resumes. If we look downstream to hiring outcomes, in Appendix Table A16, we interact the same groups with the treatment and look at their effect on an indicator for whether or not they were hired. While non-native

English speakers' writing might benefit more from the treatment[15], it does not translate into more hires relative to native English speakers. In fact, we actually see positive point estimates for effects to hiring for US and Anglophone workers, although these interaction effects are not significant. This may appear surprising, but it is important to remember that those workers are much more likely to be hired to begin with. Absent the treatment, the average worker from an Anglophone country is about twice as likely to ever be hired within their first 28 days on the platform. Because of this, in percentage terms, the treatment effect is actually larger for non-anglophone workers, 8.4%, than it is for anglophone workers whose treatment effect is 7.35%. These are not statistically different from each other, and both fit comfortably inside the 95% confidence interval on the hiring effect which is (3%,13%).

Lastly, in Appendix Table A19 we report the same specifications but look for heterogeneity in the effects to private ratings or whether the text of the review had a positive sentiment. These results are conditional on a hire, and therefore the point estimates are generally quite imprecise and we lack the power to conclude much. We can see from Column (2) that Anglophone workers are generally higher rated and Column (3) that US workers are as well. However neither have any additional effect when interacted with the treatment.

While our results suggest robust evidence for the "clarity view", it is certainly possible that there are some types of work where ones writing ability is an important indicator of their on the job performance. We look specifically at jobseekers whose primary work is in writing in Appendix Table A15. Since workers who specialize in writing make up only 17% of the sample, the standard errors are too large to be able to very confidently say anything about the effect of the treatment to ratings. Therefore, we do not reject the possibility that the signaling view could be important in jobs where writing is an important part of the output.

4.5 Robustness checks

In our main analysis, we narrow the sample to only those jobseekers whose profiles were approved by the platform. In Appendix Table A12 we run a similar regression on the full experimental sample, but we include profile approval as a control to see if it affects the estimates. In this analysis, we find that the treatment effect on the number of hires is slightly smaller than in the analysis conditional on platform approval—conditioning the sample on only jobseekers whose profiles were approved has an estimate of 7.8% while it is 10% in the full sample. The effect on the probability of any hire is 8% in the sample of only approved jobseekers and 8% in the unconditional sample. This approach and narrowing

[15] Workers from Anglophone countries have smaller treatment effects to their writing error rate in Table A14 than their Anglophone counterparts, but they still have significant positive treatment effects.

the sample to only approved jobseekers would "block" the approval channel. In Appendix Table A13 we report the same analysis not conditioned on profile approval. None of these robustness checks change the direction or significance of any of the hiring estimates, and the slightly larger estimates in the unconditional sample are unsurprising because platform approval is a necessary condition for a jobseeker to be hired.

4.6 Direct tests of the clarity view

In order to provide evidence for our hypothesized mechanism, the clarity view, we use measures of readability from the statistics, psychology, and education literature to proxy for the clarity of the text of the resumes. In Figure 8 we report the effect of the treatment on two measures of readability. These measures are based on word length, number of syllables, and sentence length. In the first facet we use a measure of reading difficulty, the Flesch-Kincaid Grade Score (Kincaid et al., 1975). The Flesch-Kincaid Grade does not have bounds, but they roughly approximate grade levels, as in a score of 12 is text approximately at a 12th grade reading level. Higher scores imply text is more difficult to read. We see that the reading difficulty score is 1% lower in the treatment group, a small but significant effect.

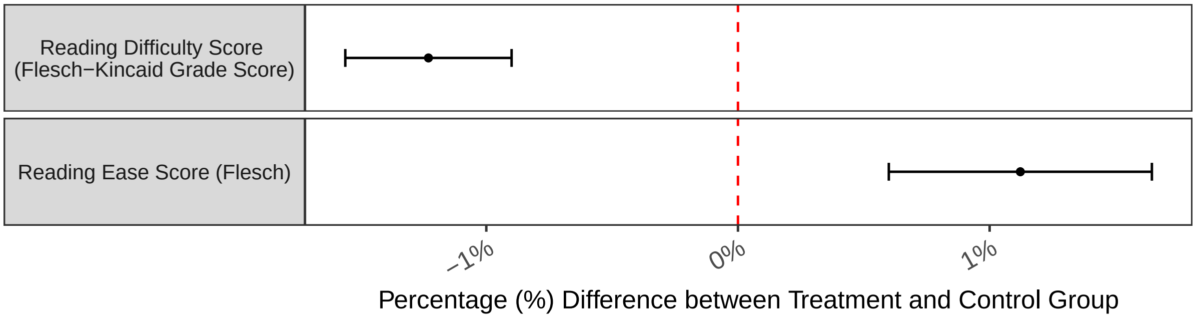
Using another measure of readability, Flesch's Reading Ease Score (Flesch, 1948), we find consistent effects— that the text of the resume's in the treatment group is easier to read. This outcome is bounded by 1 and 100, with higher scores implying easier text to read. In the control group, the reading ease score is 39.7, and 40.2 in the treatment group. While these effects are small, they consistently show that the resumes in the treatment group are easier to understand, and these measures have been used across scientific fields to understand the readability of writing (Singh et al., 2017; Alvero et al., 2021).

4.7 What happens in general equilibrium?

A first order question for platform owners and hiring managers is whether these effects would hold up as a market wide policy. As with any experimentally allocated labor market intervention, it is possible that increase in the number of workers hired does not reflect an increase in the supply or demand for labor, but instead reflect employers substituting a worker in the control group or outside of the experiment for one in the treatment group. Crowd-out concerns have been shown to be important with labor market assistance (Crépon et al., 2013).

In order to test how much of the benefits to treated workers came at the expense of other workers, in Figure 9 we break down the treatment effect by how much a jobseeker on average competed with jobseekers in the treatment group. Here create a measure of

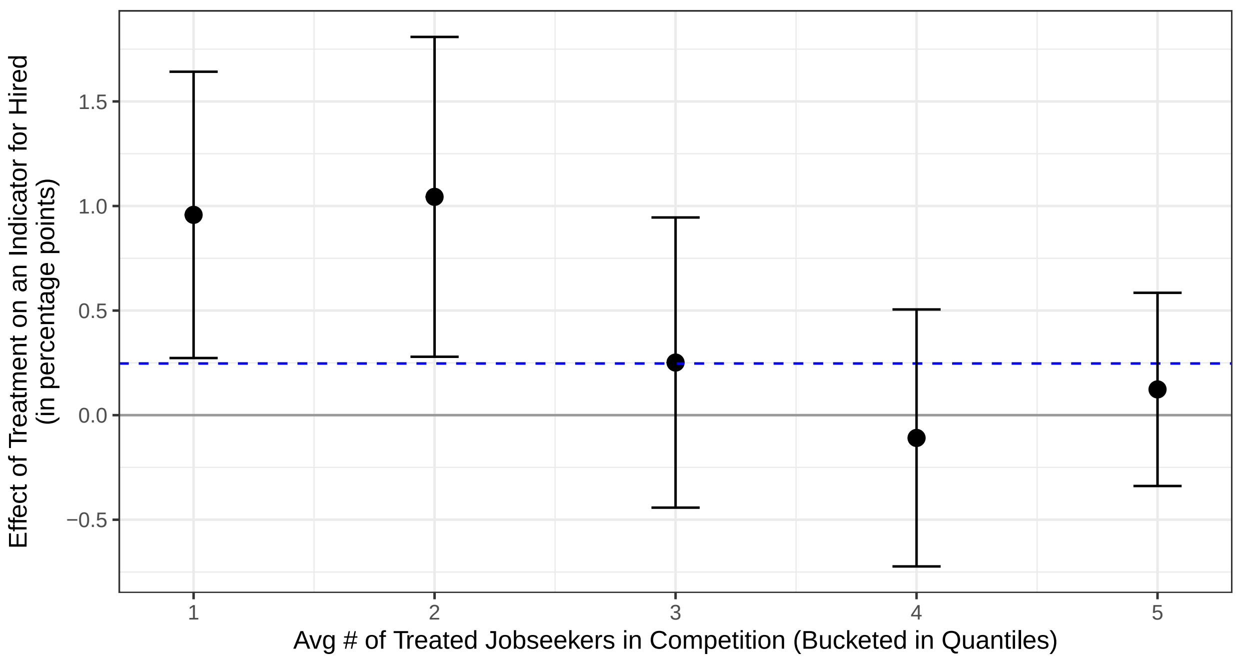
Figure 8: Effect of treatment on measures of readability



Notes: This plot shows the effect of the treatment on the readability score and grade of the profiles.The first facet plots the Reading Difficulty Score, or the Flesch-Kincaid Grade Level Score. The Flesch-Kincaid Grade does not have bounds, but they roughly approximate grade levels. Higher scores imply text is more difficult to read. In the second facet is plotted the Flesch Reading Ease Score, which is a score between 0 and 100 where the higher the score the easier it is to read. The experimental sample is of all new jobseekers who registered and were approved for the platform between June 8th and July 14th, 2021, and had non-empty resumes, with N = 194,700. See Appendix Table A20 for the regression table.

average competition with treated jobseekers. We take each job and count the number of treated jobseekers that apply. For jobseekers in the control group, we calculate the average number of treated jobseekers that apply to the jobs they apply to. To calculate the number of treated competitors for jobseekers in the treatment group, we count the number of treated jobseekers that apply to the jobs they apply to, minus one. On the x-axis of Figure 9 we break down the jobseekers into quantiles based on the average number of treated competitors they have. Jobseekers in the first quantile have one or fewer treated competitors on average, while jobseekers in the fifth quantile have more than six treated competitors on average. We find that the treatment effect diminishes based on how exposed a jobseeker is to treated jobseekers. In the first quantile, the treatment effect is a full percentage point, or more than a 30% increase in the likelihood of being hired within a month on the platform. By the third quantile, the treatment effect is almost exactly the average treatment effect of 8%, although the effect is insignificant. And for those jobseekers who compete the most with other treated jobseekers, the effect is small and insignificant.

Figure 9: Treatment Effects by Exposure to Treated Jobseekers, with ATE in blue



Notes: This analysis looks at the effect of treatment on whether or not jobseekers are hired within one month. On the x-axis we break down the jobseekers into quantiles based on the average number of treated jobseekers they compete with when they apply to jobs. The 1st quantile is jobseekers who on average apply to jobs which no more than one other treated jobseeker applied to. The 5th quantile is jobseekers who on average apply to jobs which receive 6 or more other applications from treated jobseekers. A 95% confidence interval is plotted around each estimate. The experimental sample is of all new jobseekers who registered and were approved for the platform between June 8th and July 14th, 2021, and had non-empty resumes, with N = 194,700.

5 Conclusion

Employers are more likely to hire new labor market entrants with better-written and clearer resumes. We argue that better writing makes it easier for employers to judge the match quality of a particular worker. We show results from a field experiment in an online labor market where treated workers were given algorithmic writing assistance. These jobseekers were 8% more likely to get hired and formed 7.8% more contracts over the month-long experiment. These jobseekers were hired at 10% higher wages than those in the control group, due to a change in the composition of which workers were hired. While one might have expected writing quality to be a reliable indicator of worker quality, the treatment did not affect employers' ratings of hired workers. We provide evidence for "the clarity view" of resume writing—that better writing, without any difference in the underlying facts—makes it easier for employers to correctly judge an applicants abilities.

These results have important implications for hiring managers and for platform designers. The change in the composition of hired workers to more expensive workers implies that if this technology was rolled out platform wide, it would increase platform revenue. As for the increase in number of hires, it is possible that the benefits to treated workers came at the expense of other workers, as both treated- and control-assigned workers compete in the same market. We find evidence that the treatment effect dissipates the more treated jobseekers one is in competition with. This suggests that the additional hires driven by the treatment might be crowding out other hires. However, even if additional hires came from experienced workers, this is likely still a positive result. New labor market entrants are uniquely disadvantaged (Pallais, 2014) in online labor markets. To the extent that the gains to new workers come partially at the expense of experienced workers, this is likely a good trade-off. And lastly, given the wages of the hired workers are higher with no lower ratings, when rolled out platform-wide, algorithmic writing assistance is likely to increase the quality of matches formed.

Conceptualizing AI/ML innovation and proliferation as a fall in the cost of prediction technology fits our setting (Agrawal et al., 2018a,b). Writing a resume is, in part, an applied prediction task—what combination of words and phrases, arranged in what order, are likely to maximize my pay-off from a job search? The Algorithmic Writing Service reduces the effort or cost required for making these decisions. When revising their resumes, rather than identifying errors in their own predictions themselves, jobseekers with access to the Algorithmic Writing Service are given suggestions for error correction and cleaned up writing. Furthermore, the treatment, by lowering the costs of error-free writing for at least some jobseekers, causes them to do better at writing their resumes.

These kinds of algorithmic writing assistance will likely "ruin" writing as a signal of ability. With advances in writing technologies with capabilities far beyond what is explored here (Brown et al., 2020), even if the "signaling view" was at one time dominant, the proliferation of Large Language Models are likely to make it not true in the future.

References

Abebe, Girum, A Stefano Caria, and Esteban Ortiz-Ospina, "The selection of talent: experimental and structural evidence from Ethiopia," *American Economic Review*, 2021, *111* (6), 1757–1806.

Agrawal, Ajay, John Horton, Nicola Lacetera, and Elizabeth Lyons, "Digitization and the contract labor market," *Economic Analysis of the Digital Economy*, 2015, *219.*

_ , Joshua Gans, and Avi Goldfarb, "Prediction, judgment, and complexity: a theory of decision-making and Artificial Intelligence," in "The economics of Artificial Intelligence: an agenda," University of Chicago Press, 2018, pp. 89–110.

_ , _ , and _ , *Prediction machines: the simple economics of Artificial Intelligence*, Harvard Business Press, 2018.

_ , Nicola Lacetera, and Elizabeth Lyons, "Does standardized information in online markets disproportionately benefit job applicants from less developed countries?," *Journal of international Economics*, 2016, *103*, 1–12.

Alvero, AJ, Sonia Giebel, Ben Gebre-Medhin, Anthony Lising Antonio, Mitchell L Stevens, and Benjamin W Domingue, "Essay content and style are strongly related to household income and SAT scores: Evidence from 60,000 undergraduate applications," *Science advances*, 2021, *7* (42), eabi9031.

Belot, Michèle, Philipp Kircher, and Paul Muller, "Providing advice to jobseekers at low cost: an experimental study on online advice," *The Review of Economic Studies*, 2018, *86* (4), 1411–1447.

Bertrand, Marianne and Sendhil Mullainathan, "Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination," *American Economic Review*, 2004, *94* (4), 991–1013.

Bolton, Gary, Ben Greiner, and Axel Ockenfels, "Engineering trust: reciprocity in the production of reputation information," *Management Science*, 2013, *59* (2), 265–285.

Briscese, Guglielmo, Giulio Zanella, and Veronica Quinn, "Providing government assistance online: a field experiment with the unemployed," *Journal of Policy Analysis and Management*, 2022, *41* (2), 579–602.

Brown, Tom, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell et al., "Language models are few-shot learners," *Advances in Neural Information Processing Systems*, 2020, *33*, 1877–1901.

Brynjolfsson, Erik, Danielle Li, and Lindsey R Raymond, "Generative AI at work," Technical Report, National Bureau of Economic Research 2023.

Cai, Hongbin, Ginger Zhe Jin, Chong Liu, and Li an Zhou, "Seller reputation: from word-of-mouth to centralized feedback," *International Journal of Industrial Organization*, 2014, *34*, 51–65.

Card, David, Jochen Kluve, and Andrea Weber, "Active labour market policy evaluations: A meta-analysis," *The Economic Journal*, 2010, *120* (548), F452–F477.

Chan, Jason and Jing Wang, "Hiring preferences in online labor markets: Evidence of a female hiring bias," *Management Science*, 2018, *64* (7), 2973–2994.

Crépon, Bruno, Esther Duflo, Marc Gurgand, Roland Rathelot, and Philippe Zamora, "Do labor market policies have displacement effects? Evidence from a clustered randomized experiment," *The Quarterly Journal of Economics*, 2013, *128* (2), 531–580.

Deming, David J, "The growing importance of social skills in the labor market," *The Quarterly Journal of Economics*, 2017, *132* (4), 1593–1640.

Eloundou, Tyna, Sam Manning, Pamela Mishkin, and Daniel Rock, "GPTs Are GPTs: An early look at the labor market impact potential of large language models," *arXiv preprint arXiv:2303.10130*, 2023.

Farber, Henry S, Dan Silverman, and Till Von Wachter, "Determinants of callbacks to job applications: An audit study," *American Economic Review*, 2016, *106* (5), 314–18.

Felten, Edward W, Manav Raj, and Robert Seamans, "Occupational heterogeneity in exposure to generative AI," *SSRN*, 2023.

Filippas, Apostolos, John Joseph Horton, and Joseph Golden, "Reputation inflation," *Marketing Science*, 2022.

Flesch, Rudolph, "A new readability yardstick.," *Journal of applied psychology*, 1948, *32* (3), 221.

Fradkin, Andrey, Elena Grewal, and David Holtz, "Reciprocity and unveiling in two-sided reputation systems: evidence from an experiment on airbnb," *Marketing Science*, 2021, *40* (6), 1013–1029.

Ghose, Anindya and Panagiotis G Ipeirotis, "Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics," *IEEE transactions on knowledge and data engineering*, 2010, *23* (10), 1498–1512.

Goldfarb, Avi and Catherine Tucker, "Digital economics," *Journal of Economic Literature*, 2019, *57* (1), 3–43.

HF canonical model maintainers, "DistilBERT-base-uncased-finetuned-sst-2-English (Revision bfdd146)," 2022.

Hong, Yili, Jing Peng, Gordon Burtch, and Ni Huang, "Just DM me (politely): direct messaging, politeness, and hiring outcomes in online labor markets," *Information Systems Research*, 2021, *32* (3), 786–800.

Horton, John J., "Online labor markets," *Internet and Network Economics: 6th International Workshop, WINE 2010, Stanford, CA, USA, December 13-17, 2010. Proceedings*, 2010.

—, "The effects of algorithmic labor market recommendations: evidence from a field experiment," *Journal of Labor Economics*, 2017, *35* (2), 345–385.

Kang, Sonia K, Katherine A DeCelles, András Tilcsik, and Sora Jun, "Whitened résumés: race and self-presentation in the labor market," *Administrative Science Quarterly*, 2016, *61* (3), 469–502.

Kincaid, J Peter, Robert P Fishburne Jr, Richard L Rogers, and Brad S Chissom, "Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel," 1975.

Kokkodis, Marios and Sam Ransbotham, "Learning to successfully hire in online labor markets," *Management Science*, 2022.

Luca, Michael and Oren Reshef, "The effect of price on firm reputation," *Management Science*, 2021, *67* (7), 4408–4419.

Marinescu, Ioana, "The general equilibrium impacts of unemployment insurance: Evidence from a large online job board," *Journal of Public Economics*, 2017, *150*, 14–29.

— **and Ronald Wolthoff**, "Opening the black box of the matching function: the power of words," *Journal of Labor Economics*, 2020, *38* (2), 535–568.

Martin-Lacroux, Christelle and Alain Lacroux, "Do employers forgive applicants' bad spelling in résumés?," *Business and Professional Communication Quarterly*, sep 2017, *80* (3), 321–335.

Moss-Racusin, Corinne A., John F. Dovidio, Victoria L. Brescoll, Mark J. Graham, and Jo Handelsman, "Science faculty's subtle gender biases favor male students," *Proceedings of the National Academy of Sciences*, 2012, *109* (41), 16474–16479.

Noy, Shakked and Whitney Zhang, "Experimental evidence on the productivity effects of generative Artificial Intelligence," *SSRN*, 2023.

Oreopoulos, Philip, "Why do skilled immigrants struggle in the labor market? A field experiment with thirteen thousand resumes," *American Economic Journal: Economic Policy*, 2011, *3* (4), 148–71.

Pallais, Amanda, "Inefficient hiring in entry-level labor markets," *American Economic Review*, nov 2014, *104* (11), 3565–3599.

Sajjadiani, Sima, Aaron J Sojourner, John D Kammeyer-Mueller, and Elton Mykerezi, "Using machine learning to translate applicant work history into predictors of performance and turnover," *Journal of Applied Psychology*, 2019, *104* (10), 1207.

Singh, Jyoti Prakash, Seda Irani, Nripendra P Rana, Yogesh K Dwivedi, Sunil Saumya, and Pradeep Kumar Roy, "Predicting the "helpfulness" of online consumer reviews," *Journal of Business Research*, 2017, *70*, 346–355.

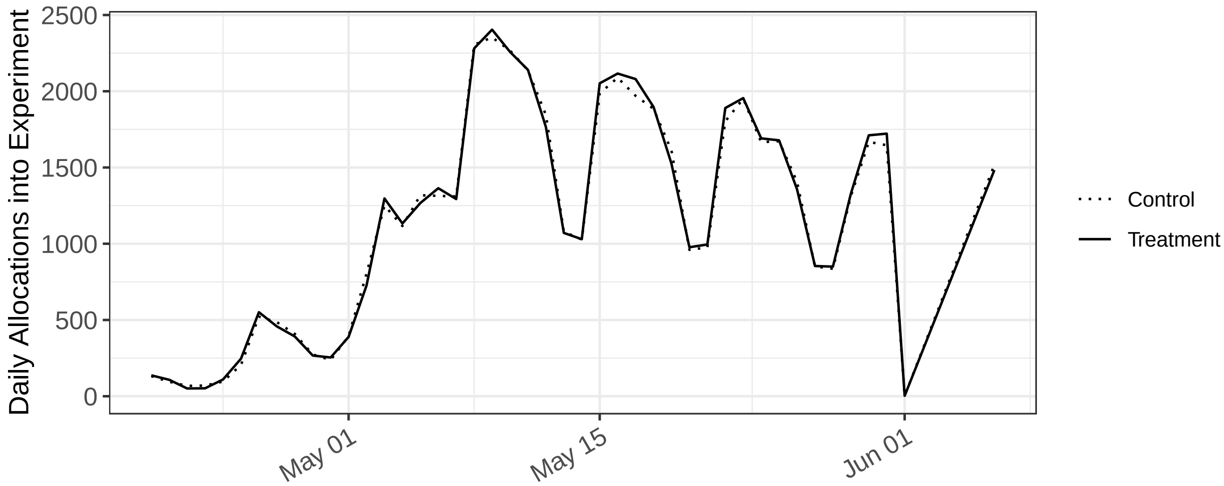
Stanton, Christopher T and Catherine Thomas, "Landing the first job: The value of intermediaries in online hiring," *The Review of Economic Studies*, 2016, *83* (2), 810–854.

Sterkens, Philippe, Ralf Caers, Marijke De Couck, Michael Geamanu, Victor Van Driessche, and Stijn Baert, "Costly mistakes: why and when spelling errors in resumes jeopardise interview chances," *Working paper*, 2021.

Tafti, Elena Ashtari, *Technology, skills, and performance: the case of robots in surgery*, Institute for Fiscal Studies, 2022.

Weiss, Daphne, Sunny X. Liu, Hannah Mieczkowski, and Jeffrey T. Hancock, "Effects of using Artificial Intelligence on interpersonal perceptions of job applicants," *Cyberpsychology, Behavior, and Social Networking*, 2022, *25* (3), 163–168.

Figure 1: Daily allocations of jobseekers into experimental cells



Notes: This plot shows the daily allocations into the treatment and control cells for the experimental sample of 480,948 new jobseekers to the platform.

A Appendix

A.1 Summary statistics and descriptives

Table A1: Summary statistics of jobs worked in the control group

	Mean	Std Deviation
Length of job in days	59	(114)
Fraction of jobs that are hourly	0.34	(0.47)
Total hours worked	201	(482)
Earnings from hourly jobs ($)	2,524	(2,524)
Earnings from fixed-price jobs ($)	386	(386)
Min hourly rate ($)	17.04	(18.42)
Max hourly rate ($)	17.22	(19)

Notes: This table reports summary statistics for jobs worked by workers in the control group of the experiment. The experimental sample is of all new jobseekers who registered and were approved for the platform between June 8th and July 14th, 2021 and had non-empty resumes. The sample of jobs include any job where the worker was hired between June 08,2021 and August 14, 2021, with N = 4,521.

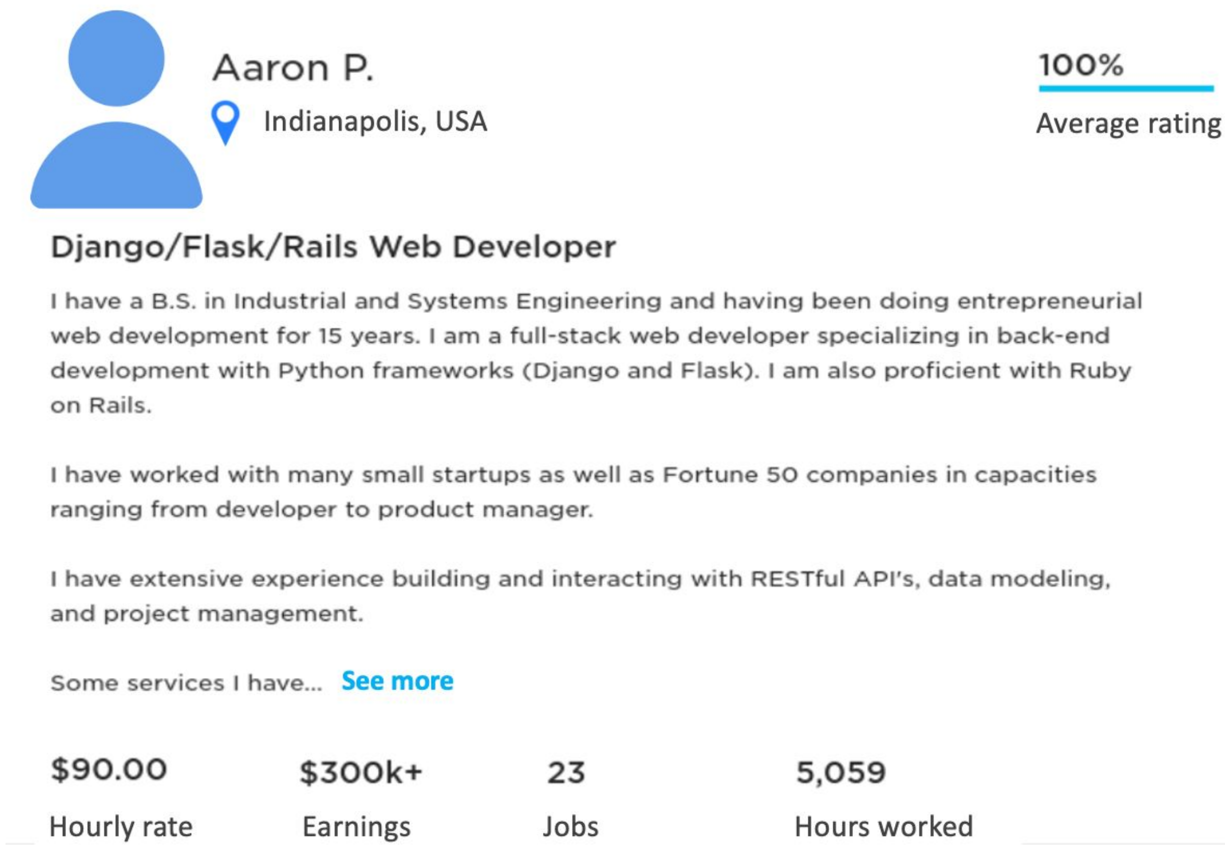


Figure 2: Stylized version of a workers' resume on the online labor market

Table A2: Summary statistics on error counts and rates in the control group

	Total Errors	Error Rate
All Error Types	4.390 (10.257)	0.080 (0.252)
Capitalization Errors	0.140 (0.543)	0.004 (0.029)
Possible Typo	2.312 (8.768)	0.041 (0.106)
Grammar Errors	0.219 (0.589)	0.004 (0.014)
Punctuation Errors	0.425 (1.629)	0.008 (0.213)
Typographic Errors	0.821 (2.985)	0.016 (0.051)
Style Errors	0.317 (0.890)	0.004 (0.011)
Miscellaneous Errors	0.103 (0.391)	0.002 (0.009)
Redundant Phrases	0.025 (0.162)	0.000 (0.003)
Nonstandard Phrases	0.002 (0.050)	0.000 (0.001)
Commonly Confused Words	0.010 (0.117)	0.000 (0.002)
Collocations	0.012 (0.121)	0.000 (0.003)
Semantic Errors	0.003 (0.061)	0.000 (0.001)

Notes: This table reports means and standard errors of the writing errors in the resumes of the control group. The first column displays the average total error count and the second column displays the average error rate (total errors normalized by word count). Writing errors are defined by LanguageToolR. The sample is made up of all jobseekers in the control group of the experimental sample who submitted non-empty resumes and were approved by the platform.

Table A3: Effects of writing assistance on profile submission and platform approval

	Dependent variable:		
	Profile submitted x 100	Approved x 100	
	(1)	(2)	(3)
Algo Writing Treatment	0.106	0.199	0.186
	(0.144)	(0.133)	(0.142)
Constant	45.532***	89.057***	40.550***
	(0.102)	(0.094)	(0.100)
Observations	480,948	219,242	480,948
R^2	0.00000	0.00001	0.00000

Notes: This table analyzes the effect of the treatment on whether or not a jobseeker's profile was submitted and approved. In Column (1) the outcome is 100 times a binary indicator for whether or not the jobseeker completed platform registration and submitted their resume, on the full experimental sample. In Column (2) the outcome is 100 times a binary indicator for whether or not the platform approved the resume, on the sample of only those jobseekers who submitted their resumes. In Column (3) the outcome is 100 times a binary indicator for whether or not the platform approved the resume, on the full experimental sample. Significance indicators: $p \leq 0.10 : *$, $p \leq 0.05 : **$ and $p \leq .01 : ***$.

Table A4: Description of Error Rule Categories with Examples

Category	Description	Examples
American English Phrases	Sentence favors the American English spelling of words.	*apologize, catalog, civilization, defense*
British English, Oxford Spelling	Sentence favors the British English spelling of words.	*apologise, catalogue, civilisation, defence*
Capitalization	Rules about detecting uppercase words where lowercase is required and vice versa.	*This house is old. it was built in 1950.* *I really like Harry potter.*
Collocations	A collocation is made up of two or more words that are commonly used together in English. This refers to an error in this type of phrase.	*Undoubtedly, this is the result of an extremely dynamic development of Lublin in the recent years.* *I will take it in to account.* *It's batter to be save then sorry.*
Commonly Confused Words	Words that are easily confused, like 'there' and 'their' in English.	*I have my won bed.* *Their elicit behavior got the students kicked out of school.It's the worse possible outcome.*
Grammar	Violations related to system of rules that allow us to structure sentences.	*Tom make his life worse.* *A study like this one rely on historical and present data.This is best way of dealing with errors.*
Miscellaneous	Miscellaneous rules that do not fit elsewhere.	*This is best way of dealing with errors.* *The train arrived a hour ago.* *It's nice, but it doesn't work. (inconsistent apostrophes)*
Nonstandard Phrases		*I never have been to London.* *List the names in an alphabetical order.* *Why would a man all of the sudden send flowers?*
Possible Typo	Spelling issues.	*It'a extremely helpful when it comes to homework.* *We haven't earned anything.This is not a HIPPA violation.*
Punctuation	Error in the marks, such as period, comma, and parentheses, used in writing to separate sentences and their elements and to clarify meaning.	*"I'm over here, she said.* *Huh I thought it was done already.* *The U.S.A is one of the largest countries.*
Redundant Phrases	Redundant phrases contain words that say the same thing twice. When one of the words is removed, the sentence still makes sense. Sometimes the sentence has to be slightly restructured, but the message remains the same.	*We have more than 100+ customers.* *He did it in a terrible way.* *The money is sufficient enough to buy the sweater.*
Semantics	Logic, content, and consistency problems.	*It allows us to both grow, focus, and flourish.* *On October 7, 2025 , we visited the client.This was my 11nd try.*
Style	General style issues not covered by other categories, like overly verbose wording.	*Moreover, the street is almost entirely residential.* *Moreover, it was named after a poet.* *Doing it this way is more easy than the previous method.* *I'm not very experienced too.* *Anyways, I don't like it.*
Typography	Problems like incorrectly used dash or quote characters.	*This is a sentence with two consecutive spaces.* *I have 3dogs.The price rose by $12,50.* *I'll buy a new T—shirt.*

Table A5: Hiring outcomes predicted based on language errors (normalized by word count) in observational data

	Dependent variable:			
	Hired	Number of Contracts	Hired	Number of Contracts
	(1)	(2)	(3)	(4)
Capitalization Error	−0.038	−0.075	−0.026	−0.055
	(0.025)	(0.048)	(0.023)	(0.045)
Possible Typo	−0.022***	−0.030**	−0.013**	−0.016
	(0.007)	(0.013)	(0.006)	(0.012)
Grammar Error	−0.314***	−0.534***	−0.210***	−0.360***
	(0.051)	(0.097)	(0.047)	(0.092)
Punctuation Error	0.0002	−0.0001	0.0001	−0.0002
	(0.003)	(0.006)	(0.003)	(0.006)
Typography Error	−0.069***	−0.098***	−0.050***	−0.066***
	(0.014)	(0.026)	(0.013)	(0.025)
Style Error	0.130**	0.261**	0.115**	0.234**
	(0.062)	(0.119)	(0.058)	(0.112)
Miscellaneous Error	−0.220***	−0.414***	−0.121	−0.252*
	(0.079)	(0.151)	(0.074)	(0.143)
Redundant Phrases	−0.264	−0.433	−0.149	−0.240
	(0.229)	(0.437)	(0.213)	(0.414)
Nonstandard Phrases	−0.124	0.804	−0.193	0.699
	(0.882)	(1.681)	(0.819)	(1.591)
Commonly Confused Words	−0.331	−0.761	−0.190	−0.531
	(0.324)	(0.618)	(0.301)	(0.584)
Collocations	−0.380*	−0.637	−0.262	−0.438
	(0.228)	(0.434)	(0.211)	(0.411)
Semantic Error	−0.532	−0.340	−0.445	−0.191
	(0.583)	(1.112)	(0.541)	(1.052)
Constant	0.036***	0.053***	0.026***	0.036***
	(0.001)	(0.002)	(0.001)	(0.002)
Controls			X	X
Observations	65,114	65,114	65,114	65,114
R^2	0.002	0.001	0.140	0.106

Notes: This table analyzes correlation between various writing errors on jobseekers' resumes and their hiring outcomes. The independent variables, writing errors, are divided by the number of words in the jobseekers' resume. Hired is defined as 1 if the jobseeker was ever hired in their first month after registering for the platform, and 0 if else. Number of Contracts is defined as the number of unique jobs they begin working in that time. Columns (3) and (4) include controls for profile hourly rate and job category. Writing errors are defined by LanguageToolR. The sample is made up of all jobseekers who registered for the platform in the week before the experiment who submitted non-empty resumes.
Significance indicators: $p \leq 0.10 : *$, $p \leq 0.05 : **$ and $p \leq .01 : ***$.

Table A6: Hiring outcomes predicted based on language errors in observational data

	Dependent variable:							
	Hired		Number of Contracts		Hired		Number of Contracts	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Total errors	-0.0002^{**}		-0.0003^{**}		-0.0001		-0.0002	
	(0.0001)		(0.0001)		(0.0001)		(0.0001)	
Num words	0.0002^{***}		0.0004^{***}		0.0001^{***}		0.0003^{***}	
	(0.00001)		(0.00002)		(0.00001)		(0.00002)	
Error rate		-0.011^{***}		-0.017^{***}		-0.007^{***}		-0.010^{**}
		(0.003)		(0.005)		(0.003)		(0.005)
Constant	0.017^{***}	0.033^{***}	0.020^{***}	0.049^{***}	0.014^{***}	0.024^{***}	0.014^{***}	0.034^{***}
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
Normalized		X		X		X		X
Controls					X	X	X	X
Observations	65,114	65,114	65,114	65,114	65,114	65,114	65,114	65,114
R^2	0.007	0.0002	0.007	0.0002	0.142	0.139	0.108	0.105

Notes: This table analyzes correlation between all writing errors on jobseekers' resumes and their hiring outcomes. The first independent variable is the total number of writing errors on a jobseekers' resume. The second independent variable is the total number of errors divided by the length of their resume, in number of words. Hired is defined as 1 if the jobseeker was ever hired in their first month after registering for the platform, and 0 if else. Number of Contracts is defined as the number of unique jobs they begin working in that time. Columns (5) through (8) include controls for profile hourly rate and job category. Writing errors are defined by LanguageToolR. The sample is made up of all jobseekers who registered for the platform in the week before the experiment who submitted non-empty resumes.
Significance indicators: $p \leq 0.10 : *$, $p \leq 0.05 : **$ and $p \leq .01 : ***$.

Table A7: Hiring outcomes predicted based on language errors in the observational data

	Dependent variable:			
	Hired	Number of Contracts	Hired	Number of Contracts
	(1)	(2)	(3)	(4)
Number of words	0.0002^{***}	0.0004^{***}	0.0001^{***}	0.0003^{***}
	(0.00001)	(0.00002)	(0.00001)	(0.00002)
Capitalization Error	-0.002^{*}	-0.006^{***}	-0.001	-0.004^{*}
	(0.001)	(0.002)	(0.001)	(0.002)
Possible Typo	-0.00001	-0.00000	0.00003	0.0001
	(0.0001)	(0.0002)	(0.0001)	(0.0002)
Grammar Error	-0.007^{***}	-0.012^{***}	-0.004^{***}	-0.008^{***}
	(0.001)	(0.002)	(0.001)	(0.002)
Punctuation Error	0.001^{***}	0.002^{*}	0.001^{*}	0.0004
	(0.0004)	(0.001)	(0.0004)	(0.001)
Typography Error	-0.001^{***}	-0.001^{**}	-0.001^{***}	-0.001^{*}
	(0.0002)	(0.0005)	(0.0002)	(0.0004)
Style Error	0.003^{***}	0.006^{***}	0.003^{***}	0.006^{***}
	(0.001)	(0.002)	(0.001)	(0.001)
Miscellaneous Error	-0.007^{***}	-0.011^{***}	-0.003^{**}	-0.005
	(0.002)	(0.003)	(0.002)	(0.003)
Redundant Phrases	-0.003	-0.009	-0.001	-0.006
	(0.004)	(0.008)	(0.004)	(0.008)
Nonstandard Phrases	-0.001	0.009	-0.003	0.005
	(0.014)	(0.026)	(0.013)	(0.025)
Commonly Confused Words	-0.008	-0.021^{*}	-0.003	-0.015
	(0.006)	(0.011)	(0.006)	(0.011)
Collocations	-0.003	-0.013	-0.003	-0.013
	(0.006)	(0.011)	(0.005)	(0.010)
Semantic Error	0.004	0.035	0.002	0.031
	(0.011)	(0.021)	(0.010)	(0.020)
Constant	0.018^{***}	0.022^{***}	0.015^{***}	0.015^{***}
	(0.001)	(0.002)	(0.001)	(0.002)
Controls			X	X
Observations	65,114	65,114	65,114	65,114
R^2	0.009	0.008	0.142	0.109

Notes: This table analyzes correlation between various writing errors on jobseekers' resumes and their hiring outcomes. Hired is defined as 1 if the jobseeker was ever hired in their first month after registering for the platform, and 0 if else. Number of Contracts is defined as the number of unique jobs they begin working in that time. Columns (3) and (4) include controls for profile hourly rate and job category. Writing errors are defined by LanguageToolR. The sample is made up of all jobseekers who registered for the platform in the week before the experiment who submitted non-empty resumes.
Significance indicators: $p \le 0.10 : *$, $p \le 0.05 : **$ and $p \le .01 : ***$.

Table A8: Effects of writing assistance on length of resume

	Dependent variable:
	Number of words in resume
Algo Writing Treatment	0.127
	(0.314)
Constant	70.541^{***}
	(0.223)
Observations	194,700
R^2	0.00000

Notes: This table analyzes the effect of the treatment on the number of words in a jobseeker's resume. The sample is made up of all jobseekers in the experimental sample who submitted non-empty profiles and were approved by the platform. Significance indicators: $p \leq 0.10 : *$, $p \leq 0.05 : **$ and $p \leq .01 : ***$.

Table A9: Effect of Treatment on Writing Errors, Page 1

	Dependent variable:					
	Capitalization	Possible Typo	Grammar	Punctuation	Typography	Style
	(1)	(2)	(3)	(4)	(5)	(6)
Algo Writing Treatment	-0.0005^{***}	-0.002^{***}	-0.0005^{***}	-0.0004	-0.002^{***}	0.0003^{***}
	(0.0001)	(0.0005)	(0.0001)	(0.0004)	(0.0003)	(0.0001)
Constant	0.003^{***}	0.041^{***}	0.004^{***}	0.010^{***}	0.015^{***}	0.004^{***}
	(0.00005)	(0.0003)	(0.00004)	(0.0003)	(0.0002)	(0.00004)
Observations	194,700	194,700	194,700	194,700	194,700	194,700
R^2	0.0003	0.0001	0.0004	0.00000	0.0004	0.0001

Notes: This table analyzes the effect of the treatment on all types of writing errors, normalized by resume length. Writing errors are defined by LanguageToolR, and divided by the number of words in a jobseekers' resume to calculate their error rate. The sample is made up of all jobseekers in the experimental sample who completed the platform registration page and submitted non-empty resume. Significance indicators: $p \leq 0.10 : *$, $p \leq 0.05 : **$ and $p \leq .01 : ***$.

Table A10: Effect of Treatment on Writing Errors, Page 2

	Dependent variable:					
	Redundant Phrases	Nonstandard Phrases	Commonly Confused Words	Collocations	Semantics	get(vars[12])
	(1)	(2)	(3)	(4)	(5)	(6)
Algo Writing Treatment	-0.0004^{***}	-0.00003^{*}	-0.00001	-0.00005^{***}	-0.0001^{***}	0.00001
	(0.00003)	(0.00001)	(0.00000)	(0.00001)	(0.00001)	(0.00001)
Constant	0.002^{***}	0.0004^{***}	0.00003^{***}	0.0001^{***}	0.0003^{***}	0.0001^{***}
	(0.00002)	(0.00001)	(0.00000)	(0.00001)	(0.00001)	(0.00000)
Observations	194,700	194,700	194,700	194,700	194,700	194,700
R^2	0.001	0.00002	0.00001	0.0001	0.0001	0.00000

Notes: This table analyzes the effect of the treatment on all types of writing errors, normalized by resume length. Writing errors are defined by LanguageToolR, and divided by the number of words in a jobseekers' resume to calculate their error rate. The sample is made up of all jobseekers in the experimental sample who completed the platform registration page and submitted non-empty resume. Significance: $p \le 0.10 : \dagger$, $p \le 0.05 : *$, $p \le 0.01 : **$ and $p \le .001 : ***$.

Table A11: Bonferroni Corrected Standard Errors for Figure 4

	Unadjusted SE	Bonferroni SE
Capitalization Errors	0.000	0.000
Possible Typo	0.0003	0.0034
Grammar Errors	0.000	0.000
Typographic Errors	0.000	0.000
Style Errors	0.000	0.000
Miscellaneous Errors	0.000	0.000
Redundant Phrases	0.056	0.556
Collocations	0.00002	0.00020
Commonly Confused Words	0.000	0.000
Semantic Errors	0.442	1.000

Table A12: Effect of algorithmic writing assistance on hiring outcomes, controlling for platform approval

	Dependent variable:				
	Num Invitations	Num Shortlists	Num Interviews	Hired x 100	Num Contracts
	(1)	(2)	(3)	(4)	(5)
Algo Writing Treatment	0.0004	0.001	0.002	0.100^{***}	0.001^{**}
	(0.001)	(0.001)	(0.002)	(0.032)	(0.001)
Approved by Platform	0.142^{***}	0.040^{***}	0.179^{***}	3.204^{***}	0.049^{***}
	(0.001)	(0.001)	(0.002)	(0.033)	(0.001)
Constant	0.00001	−0.0003	−0.001	-0.050^{*}	−0.001
	(0.001)	(0.0005)	(0.001)	(0.027)	(0.001)
Observations	480,948	480,948	480,948	480,948	480,948
R^2	0.023	0.010	0.024	0.019	0.011

Notes: This analysis looks at the effect of treatment on hiring outcomes on jobseekers in the experimental sample. The Column (1) outcome Invitations is the number of times they were recruited to a job over their first month. Column (2) is the number of times their application was shortlisted over that month. Column (3) is the number of interviews they gave over that month. Column (4) defines Hired x 100 as one hundred times the probability the jobseeker was hired over that month. Column (5) defines Number of Contracts as the number of unique jobs they work over the month after they register for the platform.

The sample used in this analysis is the entire experimental sample. Significance indicators: $p \leq 0.10 : *$, $p \leq 0.05 : **$ and $p \leq .01 : ***$.

Table A13: Effect of algorithmic writing assistance on hiring outcomes, unconditional on platform approval

	Dependent variable:				
	Num Invitations	Num Shortlists	Num Interviews	Hired x 100	Num Contracts
	(1)	(2)	(3)	(4)	(5)
Algo Writing Treatment	0.001	0.001	0.002	0.106^{***}	0.002^{**}
	(0.001)	(0.001)	(0.002)	(0.033)	(0.001)
Constant	0.058^{***}	0.016^{***}	0.072^{***}	1.249^{***}	0.019^{***}
	(0.001)	(0.0004)	(0.001)	(0.023)	(0.0005)
Observations	480,948	480,948	480,948	480,948	480,948
R^2	0.00000	0.00001	0.00000	0.00002	0.00001

Notes: This analysis looks at the effect of treatment on hiring outcomes on jobseekers in the experimental sample. The Column (1) outcome Invitations is the number of times they were recruited to a job over their first month. Column (2) is the number of times their application was shortlisted over that month. Column (3) is the number of interviews they gave over that month. Column (4) defines Hired x 100 as one hundred times the probability the jobseeker was hired over that month. Column (5) defines Number of Contracts as the number of unique jobs they work over the month after they register for the platform. The sample used in this analysis is the entire experimental sample. Significance indicators: $p \leq 0.10 : *$, $p \leq 0.05 : **$ and $p \leq .01 : ***$.

Table A14: Effects of writing assistance on error rate, by sub-groups

	Dependent variable:		
	Total Error rate x 100		
	(1)	(2)	(3)
Algo Writing Treatment	-0.512^{***}	-0.628^{***}	-0.594^{***}
	(0.070)	(0.077)	(0.075)
Anglophone Country		-4.555^{***}	
		(0.127)	
Trt ×Anglo		0.580^{***}	
		(0.179)	
US			-4.469^{***}
			(0.141)
Trt × US			0.516^{***}
			(0.200)
Constant	7.680^{***}	8.531^{***}	8.321^{***}
	(0.050)	(0.055)	(0.053)
Observations	194,700	194,700	194,700
R^2	0.0003	0.012	0.009

Notes: In Column (1) we show the overall effect of the treatment to one minus the number of errors on a jobseekers' resume divided by the number of words. In Column (2) we interact the treatment with a dummy variable for if the jobseeker is from the US, UK, Canada, or Australia. In Column (3) we interact the treatment with a dummy for if the jobseeker is in the US. The experimental sample is of all new jobseekers who registered and were approved by the platform between June 8th and July 14th, 2021 and had non-empty resumes. Significance indicators: $p \leq 0.10: *$, $p \leq 0.05: **$ and $p \leq .01: ***$.

Table A15: Effect of algorithmic writing assistance on writers

	Dependent variable:			
	Error Rate X 100	Hires X 100	Private rating	Positive text review
	(1)	(2)	(3)	(4)
Algo Writing Treatment	-0.580^{***}	0.190	-0.053	-0.064
	(0.139)	(0.184)	(0.214)	(0.060)
Constant	7.086^{***}	2.855^{***}	8.456^{***}	0.874^{***}
	(0.098)	(0.130)	(0.153)	(0.046)
Observations	33,907	33,907	672	149
R^2	0.001	0.00003	0.0001	0.008

Notes: This analysis looks at the effect of treatment on contract outcomes for jobseekers in the experimental sample whose primary job category is listed as Writing. In Column (1) the outcome is 100 times the error rate based on any error type in their resume. In Column (2) the outcome is 100 times whether or not the jobseeker ever is hired over their first month on the platform. In Column (3) the outcome is the mean private rating of jobseeker for any jobs they work in their first month on the platfrom. In Column (4) we take the text of the reviews left by employers on each job and use sentiment analysis (model: distilbert-base-uncased-finetuned-sst-2-english) to impute whether the review is positive, neutral, or negative, labeled one if it is positive or neutral. The outcome is the mean of these ratings over all contracts in the sample. The experimental sample is of all new jobseekers who registered and were approved for the platform between June 8th and July 14th, 2021 and had non-empty resumes, for all jobs they worked within 28 days of registering for the platform. Significance indicators: $p \leq 0.10 : *$, $p \leq 0.05 : **$ and $p \leq .01 : ***$.

Table A16: Effects of writing assistance on hiring, by sub-groups

	Dependent variable:		
	Hired x 100		
	(1)	(2)	(3)
Algo Writing Treatment	0.247***	0.218**	0.242***
	(0.080)	(0.088)	(0.086)
Anglophone Country		2.486***	
		(0.145)	
Trt ×Anglo		0.187	
		(0.205)	
US			2.602***
			(0.161)
Trt × US			0.072
			(0.228)
Constant	3.093***	2.629***	2.719***
	(0.057)	(0.063)	(0.061)
Observations	194,700	194,700	194,700
R^2	0.00005	0.003	0.003

Notes: This table analyzes the effect of the treatment on whether or not a jobseeker was ever hired on the platform in the month after they joined, times 100. In Column (1) we show the overall effect of the treatment to hiring. In Column (2) we interact the treatment with a dummy variable for if the jobseeker is from the US, UK, Canada, or Australia. In Column (3) we interact the treatment with a dummy for if the jobseeker is in the US. The experimental sample is of all new jobseekers who registered and were approved by the platform between June 8th and July 14th, 2021 and had non-empty resumes. Significance indicators: $p \leq 0.10 : *$, $p \leq 0.05 : **$ and $p \leq .01 : ***$.

A.2 Effect of the treatment on workers' earnings

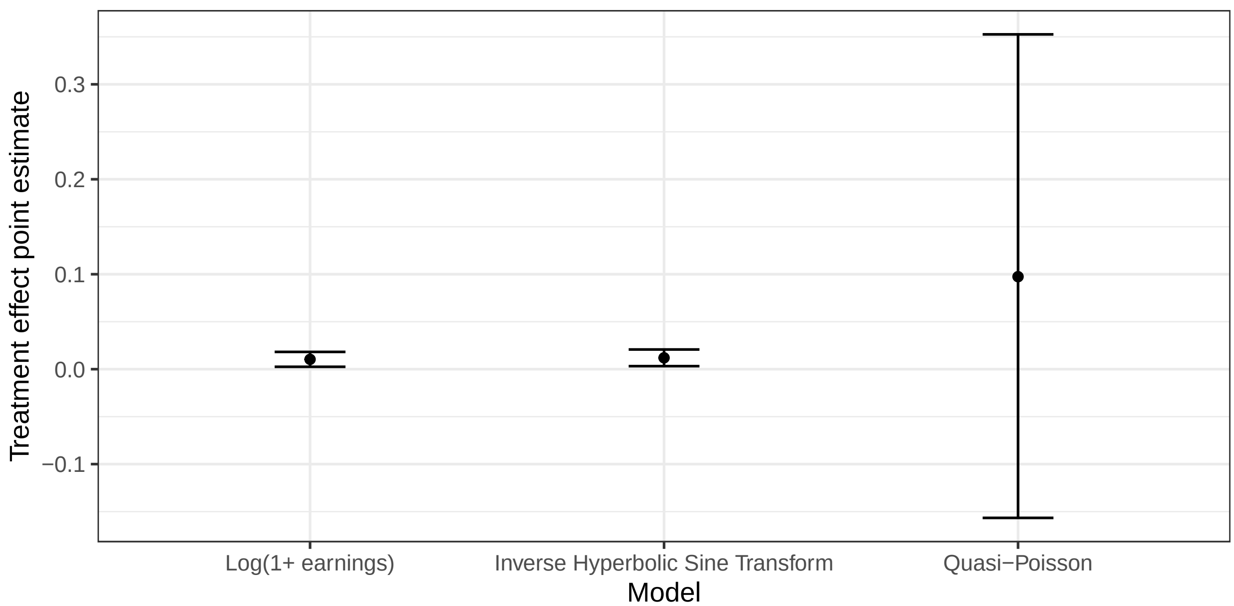
As further robustness to test whether workers in the treatment group underperformed to expectation, we regress treatment on various measures of worker's earnings in their first 28 days on the platform. In Appendix Table A17 Column (2) the outcome is workers' log of 1 plus hourly earnings. As with total hours, there are a lot of zeros, but among workers who have any hourly earnings, the average in the control group is \$1,529 in their first month. In Column (3) the outcome is the log of 1 plus the sum of workers earnings from both hourly and fixed price jobs, unconditional on ever being hired. Among workers who have any earnings, the average in the control group is \$2,957. Small positive effects on hourly earnings are mechanical due to the increase in hourly wages, and there does not seem to be any additional earnings effect to fixed price jobs. Because of the large number of jobseekers who have zero earnings, we report results from specifications which deal better with overdispersion in Appendix Figure 3. The treatment effect estimates are all small and positive, but point estimates are sensitive to specification.

Table A17: Effects of algorithmic writing assistance on workers' earnings

	Dependent variable:	
	Log hourly earnings	Log total earnings
	(1)	(2)
Algo Writing Treatment	0.010^{***}	0.010^{***}
	(0.003)	(0.004)
Constant	0.061^{***}	0.134^{***}
	(0.002)	(0.003)
Observations	194,700	194,700
R^2	0.0001	0.00003

Notes: This table analyzes the effect of the treatment on measures of workers' earnings. In Column (1) the outcome is the log hourly earnings a worker is paid over this time period. In Column (2) the outcome is the log total earnings a worker is paid over this time period. The experimental sample is of all new jobseekers who registered and were approved by the platform between June 8th and July 14th, 2021 and had non-empty resumes. Significance indicators: $p \le 0.10 : *$, $p \le 0.05 : **$ and $p \le .01 : ***$.

Figure 3: Robustness tests for total earnings treatment effect



Notes: This plot shows the effect of the treatment on the total earnings on all contracts from the workers first month on the platform. We report results from specifications meant to deal with the fact that most of the observations are zeros. The first specification is an OLS regression where the outcome is log(1+ earnings), the second specification is OLS where the earnings variable has had an Inverse Hyperbolic Sine transformation. The third is a quasi-poisson generalized linear model. The experimental sample is of all new jobseekers who registered and were approved for the platform between June 8th and July 14th, 2021, and had non-empty resumes, with N = 194,700.

Table A18: Effect of algorithmic writing assistance on workers' public ratings

	Dependent variable:					
	Communication	Skills	Quality	Cooperation	Deadlines	Recieved public rating
	(1)	(2)	(3)	(4)	(5)	(6)
Algo Writing Treatment	−0.008	−0.005	0.004	−0.008	−0.009	0.002
	(0.021)	(0.021)	(0.023)	(0.020)	(0.022)	(0.012)
Constant	4.811***	4.801***	4.768***	4.840***	4.803***	0.534***
	(0.015)	(0.015)	(0.016)	(0.015)	(0.016)	(0.009)
Observations	3,745	3,745	3,745	3,745	3,745	6,263
R^2	0.00004	0.00002	0.00001	0.00004	0.00004	0.00001

Notes: This analysis looks at the effect of treatment on the average public ratings of contracts for jobseekers in the experimental sample. All ratings are on a scale of 1 to 5 and averaged at the worker level. In Column (1) the outcome is the rating the employer gives to the workers' communication ability. In Column (2) the outcome is the rating the employer gives to the workers' skills. In Column (3) the outcome is the rating the employer gives to the overall quality of work completed. In Column (4) the outcome is the rating the employer gives to the workers' cooperation. In Column (5) the outcome is the rating the employer gives to the workers' ability to make deadlines. In Column (6) the outcome is the percentage of jobs worked where the worker was left any public rating by their employer. The experimental sample is of all new jobseekers who registered and were approved for the platform between June 8th and July 14th, 2021 and had non-empty resumes. Significance indicators: $p \leq 0.10 : *$, $p \leq 0.05 : **$ and $p \leq .01 : ***$.

Table A19: Effects of writing assistance on private ratings and reviews

	Dependent variable:					
	Private rating			Positive text review		
	(1)	(2)	(3)	(4)	(5)	(6)
Algo Writing Treatment	−0.076	−0.132	−0.126	0.016	0.015	0.014
	(0.082)	(0.098)	(0.094)	(0.019)	(0.022)	(0.021)
Anglophone Country		0.481***			0.014	
		(0.126)			(0.033)	
Trt ×Anglo		0.213			0.003	
		(0.177)			(0.044)	
US			0.516***			−0.002
			(0.135)			(0.036)
Trt × US			0.244			0.009
			(0.190)			(0.049)
Constant	8.631***	8.480***	8.502***	0.858***	0.854***	0.858***
	(0.059)	(0.071)	(0.068)	(0.014)	(0.016)	(0.016)
Observations	4,318	4,318	4,318	1,189	1,189	1,189
R^2	0.0002	0.011	0.011	0.001	0.001	0.001

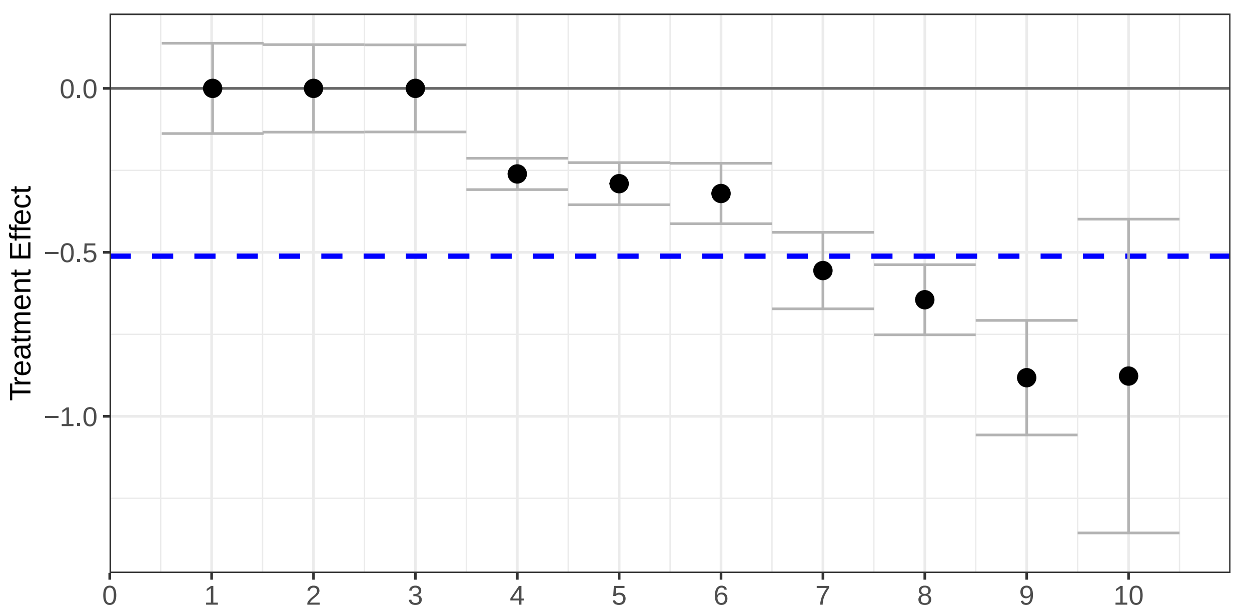
Notes: In this table we report the effect of the treatment to two measures of worker ratings. In Columns (1) through (3) the outcome is jobseekers average private ratings. In Columns (4) through (6) the outcome is the average percent of jobseekers reviews flagged as having a positive or neutral sentiment. In Column (2) and (5) we interact the treatment with a dummy variable for if the jobseeker is from the US, UK, Canada, or Australia. In Column (3) and we interact the treatment with a dummy for if the jobseeker is in the US. The experimental sample is of all new jobseekers who registered and were approved by the platform between June 8th and July 14th, 2021 and had non-empty resumes. Significance indicators: $p \leq 0.10 : *$, $p \leq 0.05 : **$ and $p \leq .01 : ***$.

Table A20: Effects of writing assistance on resume readability

	Dependent variable:	
	Reading Ease Score	Reading Difficulty Score
	(1)	(2)
Algo Writing Treatment	0.446***	−0.154***
	(0.105)	(0.021)
Constant	39.779***	12.494***
	(0.075)	(0.015)
Observations	195,247	185,487
R^2	0.0001	0.0003

Notes: This table shows the effect of the treatment on various writing readability scores. In Column (1) we show the effect of the treatment to the Flesch Reading Ease Score. This score is bounded by 1 and 100, with higher scores being more readable. In Column (2) we show the effect of the treatment to the Reading Difficulty Score, or the Flesch-Kincaid Grade Score. The score is unbounded, so we remove the most extreme 5 percent of outliers. Lower scores are more readable. The experimental sample is of all new jobseekers who registered and were approved by the platform between June 8th and July 14th, 2021 and had non-empty resumes, with outliers removed for the Flesch-Kincaid Grade Score. Significance indicators: $p \leq 0.10 : *$, $p \leq 0.05 : **$ and $p \leq .01 : ***$.

Figure 4: Effect of treatment on the error rate, by deciles



Notes: This plot shows the effect of the treatment on the number of errors divided by the number of words in jobseekers' resumes, by deciles. Jobseekers in the lowest deciles have the least writing errors relative to their length, and jobseekers at the highest deciles have the most errors. The experimental sample is of all new jobseekers who registered and were approved for the platform between June 8th and July 14th, 2021, and had non-empty resumes, with $N = 194{,}700$.

B A simple model of the "clarity view" of resume writing

In this section, we formalize a rational model of how the writing intervention could (a) increase hiring but (b) not lead to worse matches. We formalize the argument that better writing allowed employers to better ascertain who was a potential match with a simple model, and show how this kind of interplay between resume quality and hiring could exist in equilibrium.

B.1 A mass of jobseekers with heterogeneous productivity

There is a unit mass of jobseekers. If hired, their productivity is θ_i. Workers are either high-type ($\theta = \theta_H$) or low-type ($\theta = \theta_L$), with $\theta_H > \theta_L$. Workers know their own type. It is common knowledge that the fraction of high types in the market is γ. All workers, if hired, are paid their expected productivity, from the employer's point of view. Hires only last one unit of time.

B.2 Jobseekers decide whether to put effort into resume-writing

Before being hired, jobseekers write resumes. Jobseekers must decide whether to put effort $e \in \{0,1\}$ into writing that resume. Effort itself is not observable. The cost of this effort is jobseekers-specific and there is a distribution of individual resume effort costs. The support of the cost distribution is $[0, \bar{c}]$. The distribution has mass everywhere and the CDF is F and PDF is f. Jobseekers who put in no effort have resume costs of 0, while those that put in effort have a cost of c_i.

Before making an offer, firms observe a signal of jobseekers' type on their resume, $R \in \{0,1\}$. With effort, a high-type jobseeker generates an $R = 1$ signal; without effort, $R = 0$. A low-type jobseeker generates $R = 0$ no matter what. There is some share of workers λ for whom it is impossible to generate $R = 1$, regardless of their type. This share of workers are hit with a random shock of "bad writing" which make them unable to write clearly in English. There are $\gamma\lambda$ high type workers who are unable to generate $R = 0$, even if they put in effort.

Clearly, low-types will never put in effort. The question is whether a high type will put in effort. The decision hinges on whether the cost of resume effort is worth the wage premium it creates. Let $w_{R=0}$ be the wage paid in equilibrium to jobseekers with $R = 0$. Note that $w_{R=1} = \theta_H$, as there is no uncertainty about a jobseeker's type if $R = 1$.

A jobseeker i who is a high-type will choose $e = 1$ if $\theta_H - w_{R=0}(c_i) > c_i$. The marginal high-type jobseeker is indifferent between putting in effort or not, and has a resume-writing cost of $\hat{c}$, where

$$\hat{c} = \theta_H - w_{R=0}(\hat{c}). \tag{3}$$

This implies that there are $F(c)\gamma(1-\lambda)$ jobseekers that choose $e = 1$. These are the high-type jobseekers with relatively low resume writing costs who aren't hit with the "bad writing" shock. The remaining $[1 - F(c)]\gamma(1-\lambda) + \gamma\lambda$ high-type jobseekers choose $e = 0$. They are pooled together with the $1-\gamma$ jobseekers that choose $e = 0$ because they are low-types.

From the employer's perspective, if they believe that the resume effort cost of the marginal high-type jobseekers is $\hat{c}$, the probability an $R = 0$ jobseekers is high-type is

$$p_H^{R=0}(\hat{c}) = \frac{1 - F(\hat{c})\gamma(1-\lambda) + \gamma\lambda}{\gamma\lambda + (1 - F(\hat{c}))\gamma(1-\lambda) + (1-\gamma)}. \tag{4}$$

The wage received by an $R = 0$ worker is

$$w_{R=0}(\hat{c}) = \theta_L + (\theta_H - \theta_L)p_H^{R=0}(\hat{c}) \tag{5}$$

When the cost of the marginal jobseeker is higher, more jobseekers find it worth choosing $e = 1$, as $F'(\hat{c}) > 0$. This leaves fewer high-types in the $R = 0$ pool, and so

$$\frac{dp_H^{R=0}}{d\hat{c}} < 0. \tag{6}$$

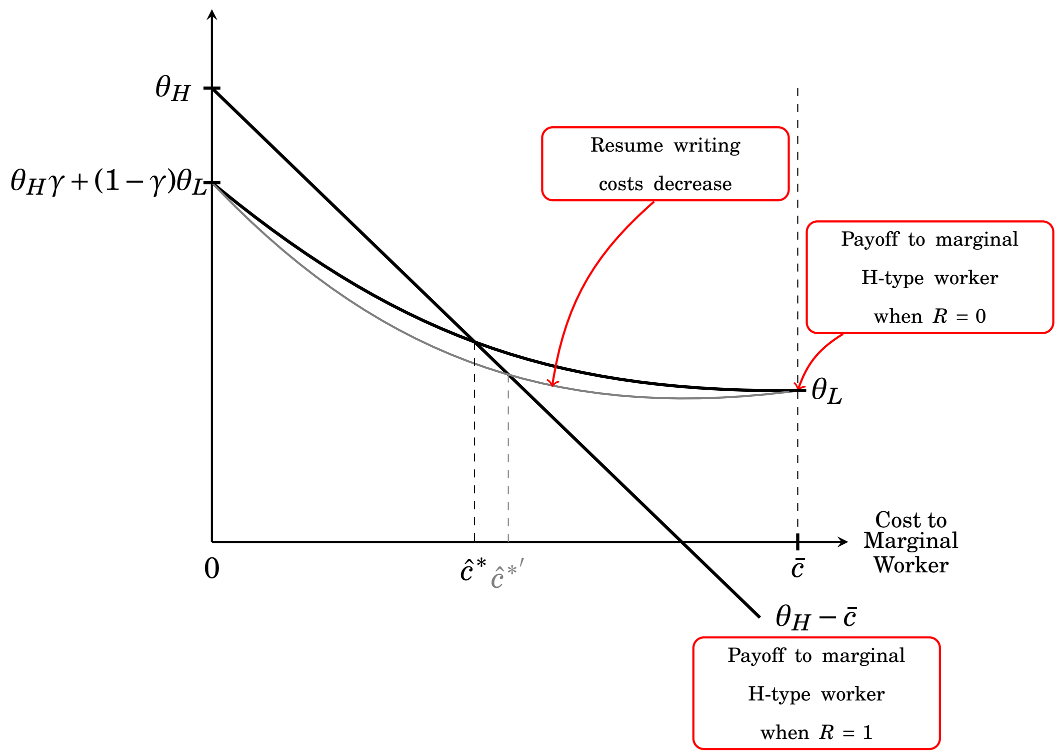
B.3 The equilibrium fraction of high-type workers putting effort into resume-writing

In equilibrium, there is some marginal high-type jobseeker indifferent between $e = 0$ and $e = 1$, and so

$$(\theta_H - \theta_L)(1 - p_H^{R=0}(\hat{c}^*)) = \hat{c}^*.$$

Figure 5 illustrates the equilibrium i.e., the cost where the marginal jobseeker is indifferent between $e = 0$ and $e = 1$. The two downward-sloping lines are the pay-offs to the marginal jobseeker for each $\hat{c}$. The pay-off to $R = 1$ is declining, as the wage is constant (at θ_H) but the cost is growing linearly. The pay-off to $R = 0$ is also declining, from Equation 6. Both curves are continuous.

Figure 5: Equilibrium determination of the marginal high-type jobseeker indifferent between putting effort into a resume



Note that when the marginal jobseeker has $\hat{c} = 0$, there is just a point-mass of high-types that have a cost that low, i.e., $f(\hat{c})$. Because the marginal jobseeker is indifferent between putting in effort and not putting in effort, jobseekers with costs of even ε will not put in effort. Since no one finds it worthwhile to put in effort the $R = 0$ pool is just the expected value of all jobseekers. And the wage is $w_{R=0}(\hat{c}) = \gamma\theta_H + (1-\gamma)\theta_L$. The marginal jobseeker pays nothing, so the pay-off is θ_H.

At the other extreme, $\hat{c} = \bar{c}$, all but a point mass of jobseekers have a cost less than this. Since the marginal jobseeker is indifferent between putting in effort at a cost of $\bar{c}$, any jobseeker with cost $\bar{c}-\varepsilon$ or below will put in effort. Then the $R = 0$ pool is purely low-types and the wage is θ_L. For the $R = 1$ market, the marginal jobseeker has a cost of $\bar{c}$ so the pay-off is $\theta_H - \bar{c}$. We know $\theta_H > \gamma\theta_H + (1-\gamma)\theta_L$. And by assumption, $\theta_L > \theta_H - \bar{c}$, and so by the intermediate value theorem, an equilibrium $\hat{c}^*$ exists on $(0, \bar{c})$.

B.4 A shift in the resume writing cost distribution leads to more high-type workers choosing to exert effort

Now suppose a technology comes along that lowers resume writing costs for some, and doesn't increase it for any, jobseekers. This technology also allows jobseekers hit with the "bad writing" shock to be able to generate $R = 1$ if they put in effort. The lower resume writing costs for those who use it shifts F higher for all points except the endpoints of the support, creating a new distribution of costs that first-order stochastically dominates the other.

Before determining the new equilibrium, note that no matter the marginal $\hat{c}$, when F increases, the probability that an $R = 0$ worker is a high-type declines, as

$$\frac{dp_H}{dF} = -\frac{1}{(F-2)^2} < 0. \tag{7}$$

This shifts the $w_{R=0}$ curve down everywhere, without changing the endpoints.

Because $w_{R=1} - \hat{c}$ is downward sloping, it intersects $w_{R=0}(\hat{c})$ at a higher value of $\hat{c}$. At the new equilibrium, the marginal jobseeker has resumes costs of $\hat{c}^{*\prime}$, where $\hat{c}^{*\prime} > \hat{c}^{*}$. At this new equilibrium, more jobseekers choose $e = 1$, causing more $R = 1$ signals. This lowers wages for the $R = 0$ group.

B.5 The effects of lower costs to welfare are theoretically ambiguous

Note that neither the shift in costs nor new found ability for λ high type workers to generate the positive signal are Pareto improving. While these high-types benefit from being able to collect $w_{R=1}$, low-types are made worse off as they find themselves in a pool with fewer high-types. Furthermore, because workers are all paid their expected product, the surplus maximizing outcome would be for everyone to choose $R = 0$. Resume effort purely changes around the allocation of the wage bill, not the total amount. Total surplus is

$$\theta_H \gamma + (1-\gamma)\theta_L - \int_0^{\bar{c}} c f(c) dc, \tag{8}$$

which is maximized at $\hat{c} = 0$, i.e., when no one finds it worthwhile to choose effort. However, with a *shift* in cost distribution (raising F), what matters is whether the marginal decrease in costs for all inframarginal workers i..e, those with $c < \hat{c}$ outweighs the costs borne by the (newly) marginal jobseekers who choose to put in effort.

In our model, all job offers are accepted. However, if we think of jobseekers as having idiosyncratic reservation values that determine whether they accept an offer, the shift in costs makes it more likely that high-types will accept an offer, while making it less likely that low-types will accept an offer. This is consistent with results where there is a greater chance an employer hires at all in the treatment. It is also consistent with our result of higher wages. Finally, if we think of employer ratings being a function of surplus, our finding of no change in employer satisfaction is also consistent, as employers are, in all cases, just paying for expected productivity.